WEST Search History

DATE: Monday, June 16, 2003

Set Name side by side	Query	Hit Count	Set Name result set
DB	*USPT,PGPB,JPAB,EPAB,DWPI; PLUR YES; OP ADJ*		
L2	(mosselman-s$.in.) AnD ((@pd > 20030205)!)	2	L2
L1	(estrogen$ near5 receptor$) AnD ((@pd > 20030205)!)	633	L1

END OF SEARCH HISTORY

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 00:24:52 ; Search time 4797.89 Seconds
(without alignments)
10988.581 Million cell updates/sec

Title: US-08-826-361A-1
Perfect score: 1434
Sequence: 1 ATGAATTACAGCATTCCCAG.....AGAACCCACAGTCTCACTGA 1434

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : GenEmbl:*
1: qb_ba:*
2: qb_htg:*
3: qb_in:*
4: qb_om:*
5: qb_ov:*
6: qb_pat:*
7: qb_ph:*
8: qb_pl:*
9: qb_pr:*
10: qb_ro:*
11: qb_sts:*
12: qb_sy:*
13: qb_un:*
14: qb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1434	100.0	1434	6	A65104	A65104 Sequence 1
2	1434	100.0	1434	6	AX323055	AX323055 Sequence
3	1434	100.0	1560	9	HSRNAERB	X99101 H.sapiens m
4	1434	100.0	1647	6	AX029400	AX029400 Sequence
5	1434	100.0	1740	9	AB006590	AB006590 Homo sapi
6	1434	100.0	1898	6	A65127	A65127 Sequence 24
7	1434	100.0	1898	6	AX323078	AX323078 Sequence
8	1434	100.0	2011	6	AX234658	AX234658 Sequence
9	1434	100.0	2011	9	AF051427	AF051427 Homo sapi
10	1432.4	99.9	1460	6	A61586	A61586 Sequence 4
11	1432.4	99.9	1460	6	AR075921	AR075921 Sequence
12	1432.4	99.9	1460	6	AX022458	AX022458 Sequence
13	1304.4	91.0	1651	9	CJERB	Y09372 C.jacchus m
14	1247	87.0	1251	6	A65105	A65105 Sequence 2
15	1247	87.0	1251	6	AX323056	AX323056 Sequence
16	1247	87.0	1257	6	A65123	A65123 Sequence 20
17	1247	87.0	1257	6	AX323074	AX323074 Sequence
18	1247	87.0	2041	9	AF051428	AF051428 Homo sapi
19	1247	87.0	2745	9	AF060555	AF060555 Homo sapi
20	1245.4	86.8	3593	9	AB006589	AB006589 Homo sapi
21	1184.6	82.6	1640	9	AF393815	AF393815 Macaca ar
22	1135.6	79.2	1584	4	AF177936	AF177936 ovis arie
23	1129.2	78.7	1611	4	AF110402	AF110402 Bos taura
24	1127.6	78.6	1820	4	BTA18017	Y18017 Bos taurus
25	1125.4	78.5	1566	9	AF393816	AF393816 Callithri
26	1106.8	77.2	1581	4	AF267736	AF267736 Sus scrof
27	1105.2	77.1	1581	4	AF164957	AF164957 Sus scrof
28	1099.6	76.7	1621	10	AF042058	AF042058 Rattus no
29	1098	76.6	1650	10	RNAJ2602	AJ002602 Rattus no
30	1096.4	76.5	2555	10	RNU57439	U57439 Rattus norv
31	1096.4	76.5	2568	6	A61583	A61583 Sequence 1
32	1096.4	76.5	2568	6	AR075920	AR075920 Sequence
33	1096.4	76.5	2568	6	AX022455	AX022455 Sequence
34	1086.8	75.8	2152	10	MMU81451	U81451 Mus musculu
35	1080.4	75.3	1458	6	A61588	A61588 Sequence 9
36	1080.4	75.3	1458	6	AR075922	AR075922 Sequence
37	1080.4	75.3	1458	6	AX022460	AX022460 Sequence
38	1080.4	75.3	1458	10	MMAJ220	AJ003220 Mus muscu
39	1035.6	72.2	1512	10	RNAJ2603	AJ002603 Rattus no
40	1035.6	72.2	1675	10	AF042059	AF042059 Rattus no
41	1032.4	72.0	1512	10	AB012721	AB012721 Rattus no
42	1022.8	71.3	1512	10	AY054413	AY054413 Mus muscu
43	959	66.9	1427	9	AF124790	AF124790 Homo sapi
44	959	66.9	2153	9	BC024181	BC024181 Homo sapi
45	878	61.2	1504	10	AF042060	AF042060 Rattus no
46	865	60.3	1215	9	AF074599	AF074599 Homo sapi
47	814	56.8	1558	10	AF042061	AF042061 Rattus no
48	811.2	56.6	840	9	AF119229	AF119229 Macaca mu
49	810.2	56.5	4163	5	AF113513	AF113513 Sturnus v
50	754.2	52.6	3539	5	AB036415	AB036415 Gallus ga
51	747.2	52.1	1419	5	AF045149	AF045149 Coturnix
52	626.4	43.7	1795	5	AF147746	AF147746 Squalus a
53	616.6	43.0	751	4	CFA313196	AJ313196 Canis fam
54	500.8	34.9	3061	5	AB003356	AB003356 Anguilla
55	486.2	33.9	1947	5	DRE414567	AJ414567 Danio rer
56	482.8	33.7	2018	5	AF349414	AF349414 Danio rer
57	466.6	32.5	2850	5	AB070630	AB070630 Paralicht
58	462.6	32.3	2283	5	AF061269	AF061269 Carassius
59	460.2	32.1	2276	5	AB070901	AB070901 oryzias l
60	457.8	31.9	1909	5	AB083064	AB083064 Cyprinus
61	453.2	31.6	2508	5	OMY289883	AJ289883 oncorhync
62	450.6	31.4	2180	5	AF136980	AF136980 Sparus au
63	449	31.3	4306	5	TNU75605	U75605 Tilapia tri
64	441.2	30.8	1772	5	AF298192	AF298192 Micropter
65	432.2	30.1	3804	5	AF185568	AF185568 Ictalurus

66	431.6	30.1	2650	5	AF177465	AF177465 Carassius
67	427	29.8	2362	5	AF298181 Micropsis	AF298181 Micropsis
68	426.2	29.7	716	4	AF257109	AF257109 ovis ari
69	417.4	29.1	1925	5	DRE414566	AJ414566 Danio rer
70	395.8	27.6	2764	6	AX376687	AX376687 Sequence
71	395.8	27.6	2770	5	ORZMER	D28954 Oryzias sp.
72	395.4	27.6	2390	5	AF298183	AF298183 Micropsis
73	395	27.5	626	10	AB076607	AB076607 Mesocricе
74	392.8	27.4	2792	5	ZFIER	S77911 Tachyglyphu
75	390.8	27.3	1374	6	AX066402	AX066402 Sequence
76	390.8	27.3	1788	6	AX411751	AX411751 Sequence
77	390.8	27.3	4963	6	AR029418	AR029418 Sequence
78	390.8	27.3	4963	6	I15368	I15368 Sequence 16
79	390.8	27.3	4963	6	156760	156760 Sequence 9
80	389.4	27.2	1728	6	E39777	E39777 Estrogen re
81	389.4	27.2	1863	6	E39778	E39778 Estrogen re
82	389.2	27.1	2092	6	AX474709	AX474709 Sequence
83	389.2	27.1	2092	6	I08538	I08538 Sequence 1
84	389.2	27.1	2092	9	HUMERMCF	M12674 Human estro
85	389.2	27.1	6450	6	AR153585	AR153585 Sequence
86	389.2	27.1	6450	9	HSERR	X03635 Human mRNA
87	389.2	27.1	6610	12	AF061181	AF061181 Mammalian
88	388.2	27.1	2085	5	GGERR	X03805 Chicken mRN
89	388	27.1	1788	4	SSESTREC	Z37167 S.scrofa mR
90	385	26.8	2552	5	DRE275911	AJ275911 Danio rer
91	383.4	26.7	5078	5	TNU75604	U75604 Tilapia nil
92	383.2	26.7	5653	4	AY033393	AY033393 ovis ari
93	382.6	26.7	2423	5	AF326201	AF326201 Halichoer
94	382.4	26.7	3416	5	AF136979	AF136979 Sparus au
95	382	26.6	1966	4	OAESTREC	Z49257 O.aries mRN
96	381.2	26.6	2821	5	AB037185	AB037185 Danio rer
97	380.8	26.6	2369	5	SAU6039	AJ006039 Sparus au
98	379.2	26.4	3500	5	AB007453	AB007453 Chrysophr
99	378	26.4	3439	5	AF349412	AF349412 Danio rer
100	375.8	26.2	1880	10	RNESTROR	X61098 Rat mRNA fo
101	375.8	26.2	2090	10	S52128	S52128 estrogen re
102	375.8	26.2	2090	10	RNERR	Y00102 Rat mRNA fo
103	375.2	26.2	1223	6	AX066401	AX066401 Sequence
104	375.2	26.2	3746	4	AF124093	AF124093 Equus cab
105	373.8	26.1	1764	5	AB055220	AB055220 Caiman cr
106	366.2	25.5	1978	10	AF181077	AF181077 Mesocricе
107	365	25.5	2296	5	AB070629	AB070629 Paralicht
108	364	25.4	2078	10	MUSESTRU	M38651 Mouse estro
109	359.8	25.1	2102	5	OCESTRE2	X84743 O.aries mRN
110	353.4	24.6	2087	5	AY055725	AY055725 Carassius
111	353.4	24.6	6041	5	AF253505	AF253505 Ictalurus
112	353.4	24.6	6756	5	AF061275	AF061275 Ictalurus
c 113	353.4	24.6	8762	5	AF253507	AF253507 Ictalurus
114	351.8	24.5	1746	5	AB055221	AB055221 Cnemidoph
115	342.2	23.9	1955	5	OMY242741	AJ242741 Oncorhync
116	342.2	23.9	2040	5	OMY242740	AJ242740 Oncorhync
117	337.4	23.5	6833	6	AX128349	AX128349 Sequence
118	337.4	23.5	6956	6	AX128348	AX128348 Sequence
119	337.4	23.5	7038	6	AX128342	AX128342 Sequence
120	336.6	23.5	1662	5	AY028456	AY028456 Petromyzo

ALIGNMENTS

```
RESULT 1
A65104
LOCUS       A65104                  1434 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 1 from Patent EP0798478.
ACCESSION   A65104
VERSION     A65104.1  GI:4530974
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1434)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798478-A 1 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     other publication CA 2200423 19970926
            other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1..1434
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       347 a    374 c    406 g    308 t
ORIGIN
```

```
  Query Match          100.0%;  Score 1434;  DB 6;  Length 1434;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-313;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACAGACAGACACTGAAAAGGAAGGTTAGTGGGAACCGTGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTGGC 240

Qy        241 GGCAGCCCTGTTACTGGTCCACGTTCAAACAGCATGCTCACTTCTGGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGATGCTCACTTCTGGCGCTGTCTGCAGC 300

Qy        301 GATTACGCATCGCGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAACAACCATTCAAGCACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAACAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434


RESULT 2
AX823055
LOCUS       AX823055               1434 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 1 from Patent EP1162264.
ACCESSION   AX823055
VERSION     AX823055.1  GI:18093942
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 1 12-DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1434
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
BASE COUNT      347 a    373 c    406 g    308 t
ORIGIN

  Query Match           100.0%;  Score 1434;  DB 6;  Length 1434;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-313;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGCC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACACAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACACAGA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
```

```
Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCACTGTATGACCTGCTGCTGGAG 1320

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCAGGGGTCCGAGTGCAGC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCAGGGGTCCGAGTGCAGC 1380

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434


RESULT 3
HSRNAERB
LOCUS       HSRNAERB                1560 bp    mRNA    linear   PRI 05-SEP-1996
DEFINITION  H.sapiens mRNA for estrogen receptor
ACCESSION   X99101
VERSION     X99101.1  GI:1518262
KEYWORDS    estrogen receptor.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1560)
  AUTHORS   Mosselman,S., Polman,J. and Dijkema,R.
  TITLE     ER beta: identification and characterization of a novel human
            estrogen receptor
  JOURNAL   FEBS Lett. 392 (1), 49-53 (1996)
  MEDLINE   96354475
   PUBMED   8769314
REFERENCE   2  (bases 1 to 1560)
  AUTHORS   Mosselman,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-1996) S. Mosselman, N.V.Organon, Biotechnology
            and Biochemistry, PO box 20 oss, Molenstraat 110 Oss, 5340 BH,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1..1560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
                     /clone_lib="Clontech; HL1010B"
                     /clone_lib="Clontech; 7414-1"
                     /dev_stage="adult"
     CDS             19..1452
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="CAA67555.1"
                     /db_xref="GI:1518263"
                     /db_xref="SPTREMBL:Q92731"
                     /translation="MNYSIDSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSH
                     LYAEPQKSPWCEARSLEHTLPVNRPTLKRKVSGNRCASPVTGPGSKRPAHFCAVCSDY
                     ASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVG
                     MVKCGSRRERCGYRLVKRQRSADEQLHCAGKAKRSGGHAPRVRELLLDAISPEQLVLT
                     LLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVR
                     LLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRF
                     RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSG
                     ISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGC
                     KSSITGSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT       367 a    408 c    445 g    340 t
ORIGIN

  Query Match           100.0%;  Score 1434;  DB 9;  Length 1560;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-413;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         19 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 78

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 138

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 198

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 258

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 318

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 378

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 438

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 GATAAAAACCGGCGCAAGAGCTGCCAGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 498

Qy        481 ATGGTGAAGPGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 558

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGCCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGCCCAAGAGAAGTGGCGGCCACGCGCCC 618

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 678

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 738

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 798

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 858

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        859 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 918

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        919 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 978

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        979 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1038

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1039 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1098

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1099 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1158
```

```
Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1159 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1218

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1219 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1278

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1279 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1338

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1339 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1398

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1399 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1452


RESULT 4
AX029400
LOCUS       AX029400               1647 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9912961.
ACCESSION   AX029400
VERSION     AX029400.1  GI:10190195
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1647)
  AUTHORS   Wilkinson,H.
  TITLE     Estrogen receptor
  JOURNAL   Patent: WO 9912961-A 2 18-MAR-1999;
            MERCK & CO INC (US) ; WILKINSON HILARY (US)
FEATURES             Location/Qualifiers
     source          1..1647
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      401 a    438 c    436 g    372 t
ORIGIN

  Query Match             100.0%;  Score 1434;  DB 6;  Length 1647;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-314;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 273

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATGGC 333

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 393

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 453

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 513

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 573

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 633

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCCACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCCACTTCGGAAGTGTTACGAAGTGGGA 693

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 753

Qy        541 AGTGCCCACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 AGTGCCCACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 813

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 873

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 933

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 993

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        994 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1053

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1113

Qy        901 CTCATCTTTGCTGCAGATCTTGTTCTGGACAGGCATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 CTCATCTTTGCTGCAGATCTTGTTCTGGACAGGCATGAGGGCAAATGCGTAGAAGGAATT 1173

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTGCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTGCAA 1233

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1293

Qy       1081 GTCACAGCGGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 GTCACAGCGGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCGTG 1353

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1354 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1413

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1473

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1533

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1593

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1594 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1647
```

```
RESULT 5
AB006590
LOCUS       AB006590                1740 bp    mRNA    linear   PRI 05-FEB-1999
DEFINITION  Homo sapiens mRNA for estrogen receptor beta, complete cds.
ACCESSION   AB006590
VERSION     AB006590.1  GI:2911151
KEYWORDS    estrogen receptor beta.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogawa,S., Inoue,S., Watanabe,T., Hiroi,H., Orimo,A., Hosoi,T.,
            Ouchi,Y. and Muramatsu,M.
  TITLE     The complete primary structure of human estrogen receptor beta (hER
            beta) and its heterodimerization with ER alpha in vivo and in vitro
  JOURNAL   Biochem. Biophys. Res. Commun. 243 (1), 122-126 (1998)
  MEDLINE   98139878
REFERENCE   2  (bases 1 to 1740)
  AUTHORS   Ogawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1997) Sumito Ogawa, Saitama Medical School,
            Department of 2nd Biochemistry; 38 Morohongo, Moroyama, Iruma-gun,
            Saitama 350-0495, Japan (E-mail:suonawa@saitama-med.ac.jp,
            Tel:81-492-76-1490, Fax:81-492-94-9751)
FEATURES             Location/Qualifiers
     source          1..1740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             99..1691
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="BAA24953.1"
                     /db_xref="GI:2911152"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT
                     GSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT      425 a    463 c    459 g    393 t
ORIGIN
```

Query Match 100.0%; Score 1434; DB 9; Length 1740;
Best Local Similarity 100.0%; Pred. No. 1.4e-314;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC  317

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC  377

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGACTCCCTGGTGTGAAGCAAGATCGCTA  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 CAGTTATCACATCTGTATGCGGAACCTCAAAAGACTCCCTGGTGTGAAGCAAGATCGCTA  437

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC  497

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC  557

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT  617

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC  677

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA  737

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA  797

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCGGGCCACGCGCCC  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCGGGCCACGCGCCC  857

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG  917

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC  977

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1037

Qy        781 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC  840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1097

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG  900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 TGTTGGATGGAGGTGTTAATGATGCGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1157

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT  960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 1217

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1277

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1337

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1397

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1457

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1517

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1577

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1637

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1691
```

```
RESULT 6
A65127
LOCUS       A65127                  1898 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 24 from Patent EP0798378.
ACCESSION   A65127
VERSION     A65127.1  GI:4530991
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1898)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 24 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     Other publication CA 2200423 19970926
            Other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      452 a    506 c    501 g    439 t
ORIGIN

  Query Match             100.0%;  Score 1434;  DB 6;  Length 1898;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-313;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCGACTTCGGAAGTGTTACGAAGTGGGA 715

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAACTCCCGCCCACCCCCCC 835

Qy        601 CCAGTGCCGGGAGCTGCTGCTGCACGCCCTGACCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CCAGTGCCGGGAGCTGCTGCTGCACGCCCTGACCCCCGAGCAGCTAGTGCTCACCCTCCTG 895

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015

Qy        781 AACAAGATTCCCGGCTTTGTGCAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 1195

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACGCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACGCTCTG 1315

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1495

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTGCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1496 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTGCCCAGTGTATGACCTGCTGCTGGAG 1555

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1556 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1615

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1616 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669


RESULT 7
AX423078
LOCUS       AX423078                1898 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 24 from Patent EP1162264.
ACCESSION   AX423078
VERSION     AX423078.1  GI:18093960
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 24 12 DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="Homo sapiens"
```

```
                     /db_xref "taxon:9606"
BASE COUNT      452 a    506 c    501 g    439 t
ORIGIN

  Query Match           100.0%;  Score 1434;  DB 6;  Length 1898;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-313;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACGCGCCC 835

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 1135

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 1195

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1495

Qy        1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1496 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1555

Qy        1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1556 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1615

Qy        1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1616 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669


RESULT 8
AX234658
LOCUS       AX234658                2011 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 2 from Patent WO0162793.
ACCESSION   AX234658
VERSION     AX234658.1  GI:15593629
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Kalush,F., Cassel,M.J., Hwang,S.S. and Winn-Deen,E.S.
  TITLE     Estrogen receptor beta variants and methods of detection thereof
  JOURNAL   Patent: WO 0162793-A 2 30-AUG-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .2011
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
BASE COUNT      470 a    541 c    561 g    439 t
ORIGIN

  Query Match           100.0%;  Score 1434;  DB 6;  Length 2011;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-313;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 637

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 697

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 757

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        758 GAACACACCTTACCTGTAAACAGAGACACACTGAAAAGCAAGGTTAGTGGGAACCGTTGC 817

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGGCTGCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGGCTGCTGCAGC 877

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 937

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 997

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1057

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGACA 1117

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGGGGCCACGCGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGGGGCCACGCGGCC 1177

Qy        601 CGACTGCCGGAGCGTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTACTGCTCACCCTGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CGACTGCCGGAGCGTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTACTGCTCACCCTGCTG 1237

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCAGTGCGCCCTTCACCGAGGCCTCC 1297

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1357

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 AAGAACATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1417

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGCCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGCCGCTCAATTGACCACCCCGGCAAG 1477

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1537

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTTCGAGAGTTAAAATCGCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTTCGAGAGTTAAAATCGCAA 1597

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1657

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1717

Qy       1141 ACCGATGCTTTGGTTTGGGCTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 ACCGATGCTTTGGTTTGGGCTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1777

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1778 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1837

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1838 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1897

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1898 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1957

Qy       1381 CGGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1958 CGGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 2011


RESULT 9
AF051427
LOCUS       AF051427                2011 bp    mRNA    linear   PRI 14-JUL-1999
DEFINITION  Homo sapiens estrogen receptor beta mRNA, complete cds.
ACCESSION   AF051427
VERSION     AF051427.1  GI:2970563
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Moore,J.T., McKee,D.D., Slentz-Kesler,K., Moore,L.B., Jones,S.A.,
            Horne,E.L., Su,J.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Cloning and characterization of human estrogen receptor beta
            isoforms
  JOURNAL   Biochem. Biophys. Res. Commun. 247 (1), 75-78 (1998)
  MEDLINE   98300286
   PUBMED   9636657
REFERENCE   2  (bases 1 to 2011)
  AUTHORS   Moore,J.T., McKee,D.D., Moore,L.B., Jones,S.A., Su,J.-L.,
            Slentz-Kesler,K., Horne,E.L., Kliewer,S.A., Lehmann,J.M. and
            Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..2011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             419..2011
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAC05985.1"
                     /db_xref="GI:2970564"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEIDGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVINLEGGPCRQLISPNVLWPTPCHLSPLVVHHQLSHLYAEP
                     QKSPWCEAPSLEHTLPVNPETLKPKVSGNRCASPVTGPGSKPDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKG
                     SRKERCGYRLVKRQKSADEQLHCAGKAKRSGGHAPEVRELLLRALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELK
                     LQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT
                     GSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT      470 a    541 c    561 g    439 t
ORIGIN

  Query Match           100.0%;  Score 1434;  DB 9;  Length 2011;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-313;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 637

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 697

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
```

```
Db        698 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 757

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Db        758 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 817

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Db        818 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 877

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Db        878 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 937

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Db        938 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 997

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Db        998 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1057

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Db       1058 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1117

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Db       1118 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 1177

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Db       1178 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 1237

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Db       1238 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 1297

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Db       1298 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1357

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Db       1358 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1417

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Db       1418 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1477

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGCAATT 960

Db       1478 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGCAATT 1537

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Db       1538 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1597

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Db       1598 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1657

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Db       1658 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1717

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Db       1718 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1777

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260

Db       1778 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1837

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320

Db       1838 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1897

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380

Db       1898 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1957

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434

Db       1958 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 2011


RESULT 10
A61586
LOCUS       A61586                  1460 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 4 from Patent WO9709348.
ACCESSION   A61586
VERSION     A61586.1  GI:3715917
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1460)
  AUTHORS   Kuiper,G.G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348-A 4 13-MAR-1997;
            KAROBIO AB (SE)
FEATURES             Location/Qualifiers
     source          1..1460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match           99.9%;  Score 1432.4;  DB 6;  Length 1460;
  Best Local Similarity   99.9%;  Pred. No. 4.3e-313;
  Matches 1433;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Db         27 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 86

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Db         87 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 146

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Db        147 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 206

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Db        207 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 266

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Db        267 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 326

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Db        327 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 386

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Db        387 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 446

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Db        447 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 506
```

```
Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGGCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGGCTTGTGCGGAGACAGAGA 566

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 626

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 686

Qy        661 GAGGCTGAGCCCCCCCATGTCTGATCACGCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 GAGGCTGAGCCCCCCCATGTCTGATCACGCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 746

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 806

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGCCTCTTCGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGCCTCTTCGAGAGC 866

Qy        841 TGTTGGATGCAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCCGCAAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TGTTGGATGCAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCCGCAAC 926

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGCGATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGCGATT 986

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAAGTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAAGTCCAA 1046

Qy       1021 CACAAACAATATCTCTGTGTCAAGGCCATGATCCTGGTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGGTCAATTCCAGTATGTACCCTCTG 1106

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAAGGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAAGGCCGTG 1166

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1226

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1286

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1287 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1346

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1347 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1406

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1407 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1460


RESULT 11
AR075921
LOCUS       AR075921               1460 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 4 from patent US 5958710.
ACCESSION   AR075921
VERSION     AR075921.1  GI:10002667
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1460)
  AUTHORS   Kolper,G., Enmark,P.I. K. and Gustafsson,J.-A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710-A 4 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1. .1460
                     /organism="unknown"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match           99.9%;  Score 1432.4;  DB 6;  Length 1460;
  Best Local Similarity   99.9%;  Prod. No. 4.3e-413;
  Matches 1433;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 86

Qy         61 ACAAGCCCAAATCTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTACTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTACTGGTCCATCGC 146

Qy        121 CAGTTATCACATCTGTATGCGCAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 CAGTTATCACATCTGTATGCGCAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 206

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 266

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 326

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 386

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 446

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 506

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGGCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGGCTTGTGCGGAGACAGAGA 566

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 626

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 686

Qy        661 GAGGCTGAGCCCCCCCATGTCTGATCACGCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 GAGGCTGAGCCCCCCCATGTCTGATCACGCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 746

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 806

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGCCTCTTCGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGCCTCTTCGAGAGC 866
```

```
Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 926

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 986

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1046

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1106

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1166

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1226

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1286

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1287 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1346

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1347 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1406

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              |||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1407 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCTACAGTCTCAGTGA 1460
```

```
RESULT 12
AX022458
LOCUS       AX022458                1460 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 4 from Patent EP0935000.
ACCESSION   AX022458
VERSION     AX022458.1  GI:10046069
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FEATURES             Location/Qualifiers
     source          1..1460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match             99.9%;  Score 1442.4;  DB 6;  Length 1460;
  Best Local Similarity   99.9%;  Pred. No. 4.3e-313;
  Matches 1433;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 86

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 146

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 206

Qy        181 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 266

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 326

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 386

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 446

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 506

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 566

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 626

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 686

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 746

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 806

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 866

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 926

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 986

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1046

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1106

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1166

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1226

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1286
```

```
Qy       1261 GAACATCTGCTCAACATCAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1287 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1346

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1347 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1406

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTCA 1434
             ||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1407 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCTACAGTCTCAGTCA 1460


RESULT 13
CJERB
LOCUS       CJERB                   1651 bp    mRNA    linear   PRI 11-NOV-1999
DEFINITION  C.jacchus mRNA for estrogen receptor beta.
ACCESSION   Y09372
VERSION     Y09372.2  GI:6433823
KEYWORDS    DNA binding; estrogen receptor; steroid hormone receptor;
            transactivation.
SOURCE      Callithrix jacchus.
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
            Callithrix.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Gaughan,J.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1651)
  AUTHORS   Gaughan,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1996) J. Gaughan, Medical Research Council,
            Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh, EH3
            9EW, UK
REFERENCE   3  (bases 1 to 1651)
  AUTHORS   Scobie,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) G. Scobie, Medical Research Council
            Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh, EH3
            9EW, UK
COMMENT     On Nov 15, 1999 this sequence version replaced gi:1669661.
            Related entries: X99101 & D57439.
FEATURES             Location/Qualifiers
     source          1..1651
                     /organism="Callithrix jacchus"
                     /db_xref="taxon:9483"
                     /clone="pCEM T vector"
                     /tissue_type="prostate"
                     /tissue_type="ovary"
                     /tissue_type="testis"
     CDS             52..1644
                     /function="steroid hormone receptor"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="CAA70546.2"
                     /db_xref="GI:6433824"
                     /db_xref="SPTREMBL:Q95171"
                     /translation="MDIKNSPSSLNSPSSYNPQQSIFSYEQSRFYGGEPTFIFSSIFMSPVASAYEGPQVKSPWCYTGNEASLFLLDPAFVGTPSWRVTTGSKRDAHFQAVCSDYASGVHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERGGYRLVRRQGNAEFQLHCAGKAKSGGHVPRVRELISAPSPFQLVITQHLPANSPPHVLISRPSVPPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWLEVLMVGLMWRSIDHPGKLIFAPNLLLDRDEGKCVEGILEVFDMLLATTSRFRELKLQHKEYLCVKAMVLLNSQYDPLVTATQDAESSQKLAHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHIRHASNKGMEHLLSMKCKNVVPVYDLLLEMMNAHVLRGCKSSITGSECSPAEDSKSIEGSQNPQSP"
BASE COUNT      397 a    434 c    443 g    377 t
ORIGIN

  Query Match             91.0%;  Score 1304.4;  DB 9;  Length 1651;
  Best Local Similarity   94.4%;  Pred. No. 3.1e-284;
  Matches 1353;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              |||||||||||||||||||| ||| ||||||||||||  ||||||||||||||||| |
Db        211 ATGCAATTACAGCATTCCCAGCAGTGTTACTAACTTGGAAGAGGGGCCTGGTCGGCAGATC 270

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATGGC 120
              |||||||||||| || |||| |||||||||| ||||||||||||||||  |||||| |
Db        271 ACAAGCCCAAATATGCTGTGGTCAACACCTGGCCACCTTTCTCCTTTAGTGTCCATCAC 330

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||
Db        331 CAGTTATCACATCTGTATGCAGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 390

Qy        181 GAACACACCTTACCTGTAAACAGACAGACACTGAAAAGCAAGGTTAGTGGGAAGCCTTG2 240
              ||||| |||||||||||| ||||||||||||||||||| |||||||||||| ||
Db        391 GAACATACCTTACCTGTAAGCAGAGACACTGAAAAGGAAAGTTAGTGGGAACCATTGT 450

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||
Db        451 GCCAGCCCTGTTACTGGTCCAAGTTCAAAGAGGGATGCTCACTTCTGTGCAGTCTGCAGC 510

Qy        301 GATTACGGCATCGGGATATCACTATGGAGTCTGGTCGTGTGCAAGGATGTAAGGCCTTTTT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GATTACGGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTT 570

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 630

Qy        421 GATAAAAACCGGCCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        631 GATAAAAACCGGCCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGCTACGAAGTGGGA 690

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        691 ATGGTAAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGGGA 750

Qy        541 AGTGCCGAGAGCAGCTGCACTGTGCCGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              | |||||| |||||||||| || |||||||||||||||||||||||||| || |||| ||||
Db        751 AATGCCGAGGAGCAGCTGCATTGCGCCGGCAAGGCCAAGAGAAGTGGTGGTCACGTGCCC 810

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              |||||||||||||||||||  ||||||||||||||||||| ||||||||||||||||||
Db        811 CGAGTGCGGGAGCTGCTGCTGAGCGCCCTGAGCCCCGAGCAGCTGGTGCTCACCCTCCTG 870

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||| || |||||||||||||||||||||||||||| | |||||||| ||||||||||
Db        871 GAGGCCGAACCGCCCCCATGTGCTGATCAGCCGCCCCAGCCTGCCCTTCACTGAGGCCTCC 930

Qy        721 ATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGTTGGTCACACATGATCAGCTGGGCT 780
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        931 ATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGTTGGTCACACATGATCAGCTGGGCT 990

Qy        781 AAGAAGATTCCCGGCTTTGTGGACCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        991 AAGAAGATTCCCGGCTTTGTGGACCTCAGCCTGTTGGACCAAGTGCGGCTCTTGGAGAGC 1050

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1051 TGCTGGTTGGAGGTGTTAATGCTGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1110

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAAATGCGTAGAAGGAATT 960
              ||||||||||||||| ||| |||||||||||||||||||||||||||||||||||||||||
Db       1111 CTCATCTTTGCTCCAAATCTCATTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 1170

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       1171 CTGGAAGTCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTGAAACTCCAA 1230

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||| |||||||||||||||||||||| |||||||||||||       ||||||||
```

```
Db        1241 CACAAAGAGTATCTCTGTGTCAAGGCCATGGTCCTGCTCAATTCCCAGTATGACCCTCTG 1290

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
               ||||||||| |||||||||||||| ||||||| |||||||| |||||||||| ||||||
Db        1291 GTCACAGCAACCCAGGATGCTGAGAGCAGCCAGAAGCTGGCCCACTTGCTGAATGCCGTG 1350

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               || || |||||||| |||||||| || ||||| ||| ||||||||||||||||| || ||
Db        1351 ACTGACGCTTTGGTCTGGGTGATCGCTAAGAGTGGCTTCTCCTCCCAGCAGCAGTCGGTG 1410

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        1411 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACATCAGGCATGCGAGTAACAAGGGCATG 1470

Qy        1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
               ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        1471 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1530

Qy        1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
               ||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        1531 ATGATGAATGCCCACGTGGTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1590

Qy        1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
               |||||||||||||||||||| |||||||||||||||||||||||||| ||||
Db        1591 CCGGCAGAGGACAGTAAAAGCACAGAGGGCTCCCAGAACCCACAGTCTCCGTGA 1644


RESULT 14
A65105
LOCUS       A65105                   1251 bp    DNA     linear   PAT 29-MAR 1999
DEFINITION  Sequence 2 from Patent EP0798378.
ACCESSION   A65105
VERSION     A65105.1  GI:4530974
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1251)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 2 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     other publication CA 2200423 19970926
            other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1..1251
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      297 a    424 c    454 g    276 t
ORIGIN

  Query Match             87.0%;  Score 1247;  DB 6;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGCAAG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGCAAG 900

Qy         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               |||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 15
AX823056
LOCUS       AX823056                 1251 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 2 from Patent EP1162264.
ACCESSION   AX823056
VERSION     AX823056.1  GI:18093943
```

```
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264 A 2 12-DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1251
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      297 a    324 c    354 g    276 t
ORIGIN

  Query Match             87.0%;  Score 1247;  DB 6;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTCCGCTGTCTCCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTCCGCTGTCTCCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTGCCCTGACCAAGTTGGCGGACAAGGAGTTGCTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTGCCCTGACCAAGTTGGCGGACAAGGAGTTGCTACACATGATCAGCTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 16
A65123
LOCUS       A65123                  1257 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 20 from Patent EP0798378.
ACCESSION   A65123
VERSION     A65123.1  GI:4530988
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1257)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 20 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     Other publication CA 2200423 19970926
            Other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1. .1257
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      297 a    327 c    355 g    278 t
ORIGIN

  Query Match             87.0%;  Score 1247;  DB 6;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
```

```
Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGACAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCCGCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
```

```
RESULT 17
AX823074
LOCUS       AX823074                1257 bp    DNA     linear   PAT 07-JAN-2004
DEFINITION  Sequence 20 from Patent EP1162264.
ACCESSION   AX823074
VERSION     AX823074.1  GI:18093957
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 20 12-DEC 2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1..1257
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      297 a    327 c    355 g    278 t
ORIGIN

  Query Match             87.0%;  Score 1247;  DB 6;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGACAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCCGCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAGGGGAAATGCCTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAGGGGAAATGCCTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCGTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATC 1200

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGGCAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGGCAG 1247


RESULT 18
AF051428
LOCUS       AF051428                2041 bp    mRNA    linear   PRI 17-MAR-1998
DEFINITION  Homo sapiens estrogen receptor beta2 splice variant mRNA, complete
            cds.
ACCESSION   AF051428
VERSION     AF051428.1  GI:2961558
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2041)
  AUTHORS   Moore,J.T., McKee,D.D., Moore,L.B., Jones,S.A., Su,J.-L.,
            Horne,E.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Cloning, Ligand Binding and Functional Activity of human estrogen
            Receptor beta Isoforms
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2041)
  AUTHORS   Moore,J.T., McKee,D.D., Moore,L.B., Jones,S.A., Su,J.-L.,
            Horne,E.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25 FEB 1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..2041
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             419..1906
                     /codon_start=1
                     /product="estrogen receptor beta2 splice variant"
                     /protein_id="AAC05751.1"
                     /db_xref="GI:2961559"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPIPGHLSPLVVHRQLSHLYAEP
                     QPSPWCEAPSLEHILPVNPETLKKEVSGNPCASPVTGGSKPMDAHPCAAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SKRERGGYRLVKRQKSADEQLHCAGKAKKSGGHAPKVKELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT
                     GSECSPAEDSKSTESPQ"
BASE COUNT      505 a    532 c    544 g    460 t
ORIGIN

  Query Match             87.0%;  Score 1247;  DB 9;  Length 2041;
  Best Local Similarity   100.0%;  Pred. No. 4e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 637

Qy         61 ACAAGCCCAAAIGTGTTGTGGCCAACACCTGGGCACGTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 ACAAGCCCAAAIGTGTTGTGGCCAACACCTGGGCACGTTTCTCCTTTAGTGGTCCATCGC 697

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGTGTGGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGGAAGCAAGATCGCTA 757

Qy        181 GAACACACCTTACGTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 GAACACACCTTACGTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 817

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 877

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGAIGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 937

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 997

Qy        421 GATAAAAACCGGCGCCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1057

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1117

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAAGTCGGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAAGTCGGGCGGCCACGCGCCC 1177

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 1237

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCC 1297

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 1357

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1417

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGGCTCAATTGACCACCCCGGCAAG 1477

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1537

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1597

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1657

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1717

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1777

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
                ||||||||||||||||||||||||||||||||||||||||||||||
Db       1778 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1824


RESULT 19
AF060555
LOCUS       AF060555                2745 bp    mRNA    linear   PRI 29-APR-1998
DEFINITION  Homo sapiens estrogen receptor beta 3 isoform mRNA, complete cds.
ACCESSION   AF060555
VERSION     AF060555.1  GI:3091285
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2745)
  AUTHORS   Moore,J.T., McKee,D.D., Slentz-Kesler,K., Moore,L.B., Jones,S.A.,
            Su,J.L., Horne,E.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..2745
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             471..2012
                     /note="splice variant"
                     /codon_start=1
                     /product="estrogen receptor beta 3 isoform"
                     /protein_id="AAC15234.1"
                     /db_xref="GI:3091286"
                     /translation="MDIKNSPSSSLNSPSSYNQSQSILPLEHQSIYIPSSYVDSHHEYP
                     AMTPYSFAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHHQLSHLYAEP
                     QKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKG
                     SRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSI
                     TGSECSPAEDSKSKEGGSPPRE"
BASE COUNT       741 a    665 c    690 g    649 t
ORIGIN

  Query Match          87.0%;  Score 1247;  DB 9;  Length 2745;
  Best Local Similarity   100.0%;  Pred. No. 3e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 689

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 749

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 809

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 869

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 929

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 989

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 1049

Qy        421 GATAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 GATAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1109

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1169

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 1229

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 1289

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 1349

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1409

Qy        781 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1469

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGGCTCAATTGACCACCCCGGCAAG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1470 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGGCTCAATTGACCACCCCGGCAAG 1529

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1589

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1649

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1709

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1710 GTCACAGCCACCCAGGATGCTCACAGCAGCCCGGAAGCTGGCTCACTTGCTCAACCCCTC 1769

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1770 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1829

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               |||||||||||||||||||||||||||||||||||||||||||||
Db        1830 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1876


RESULT 20
AB006589
LOCUS       AB006589                4593 bp    mRNA    linear   PRI 28 JUL 1998
DEFINITION  Homo sapiens mRNA for estrogen receptor beta cx, complete cds.
ACCESSION   AB006589
VERSION     AB006589.1  GI:3345671
KEYWORDS    estrogen receptor beta cx.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogawa,S., Inoue,S., Watanabe,T., Orimo,A., Hosoi,T., Ouchi,Y. and
            Muramatsu,M.
  TITLE     Molecular cloning and characterization of human estrogen receptor
            betacx: a potential inhibitor ofestrogen action in human
  JOURNAL   Nucleic Acids Res. 26 (15), 3505-3512 (1998)
  MEDLINE   98337908
REFERENCE   2  (bases 1 to 3593)
  AUTHORS   Ogawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1997) Sumito Ogawa, Saitama Medical School,
            Department of 2nd Biochemistry; 38 Morohongo, Moroyama, Iruma-gun,
            Saitama 350-0495, Japan (E-mail:suoqawa@saitama-med.ac.jp,
            Tel:81-492-76 1490, Fax:81-492-94-9751)
FEATURES             Location/Qualifiers
     source          1. .3593
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             1278. .2763
                     /codon_start=1
                     /product="estrogen receptor beta cx"
                     /protein_id "BAA31966.1"
                     /db_xref="GI:3345672"
                     /translation="MDIKNSPSSSNSPSSYNCSQSILPLEHGSIYIPSSTVDSHHETF
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVIWPTPCHLSPLVVHRQLSHIYAEP
                     QKSPWCFARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SRRERCGYRLVRRQRSADEQLHCACKAKREGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSI
                     TGSECSPAEDGGLKEPESPVTLEAVVPLYNDLLLDAEMLGSGHEKESKPFRSDSWMC
                     PREMK"
BASE COUNT       971 a    811 c    957 g    854 t
ORIGIN

  Query Match          86.8%;  Score 1245.4;  DB 9;  Length 4593;
  Best Local Similarity   99.9%;  Pred. No. 7e-271;
  Matches 1246;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 1494

Qy         61 ACAAGCCCAAATGTGTTCTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 1554

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 1614

Qy        181 GAACACACGTTACGTCTAAACAGACAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 1674

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGGGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGGGCTGTCTGCAGC 1734

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCTTTTTT 1794

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 1854

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1914

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1974

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCC 2034

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 2094

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 2154

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 2214

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCCTGGAGAGC 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTACGGCTCCTGGAGAGC 2274

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGCAAG 2334

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGCACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CTCATCTTTGCTCCAGATCTTGTTCTGCACAGGGATGAGGGCAAATGCGTAGAAGGAATT 2394

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 2454

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2455 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 2514

Qy       1081 GTCACAGCGGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2515 GTCACAGCGGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 2574

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2575 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 2634

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||
Db       2635 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 2681
```

```
RESULT 21
AF393815
LOCUS       AF393815                1640 bp    mRNA    linear   PRI 05 JUL 2001
DEFINITION  Macaca arctoides estrogen receptor beta 2 (ERBeta2) mRNA, complete
            cds.
ACCESSION   AF393815
VERSION     AF393815.1  GI:14600266
KEYWORDS    .
SOURCE      Macaca arctoides.
  ORGANISM  Macaca arctoides
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1640)
  AUTHORS   Scobie,G.A., Wilson,J.A., Millar,M.R., Macpherson,S. and
            Saunders,P.T.
  TITLE     The estrogen receptor beta variant ERBeta cx/ERBeta2 is expressed
            in a wide range of tissues in both Old and New World primates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1640)
  AUTHORS   Scobie,G.A. and Saunders,P.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21 JUN 2001) Medical Research Council, Human
            Reproductive Sciences Unit, 47 Chalmers Street, Edinburgh EH3 9ET,
            UK
FEATURES             Location/Qualifiers
     source          1. .1640
                     /organism="Macaca arctoides"
                     /db_xref="taxon:9540"
     gene            1. .1640
                     /gene="ERBeta2"
     CDS             98. .1537
                     /gene="ERBeta2"
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta 2"
                     /protein_id="AAK71317.1"
                     /db_xref="GI:14600267"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVESHHEYP
                     AMTFYSPAVMNYSIPSNVTNIEGGPGRQPASPNVLWPTPCHESPLAVHPQISHLYAEP
                     QKSPWCEARSLEHTLPVNKETLKRKVSGNKQTSPVTSPSSKRDAHFCAVCSDFASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SKRERCGYRLVRRQRSADFQVHCASKAKRSGSHTPLVRELLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSNMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSI
                     TGSECSPAEDSKSKEGGSPPKPLKKLLQREKDTSAGDKEMDKVACPGGASQNPNEKW
                     YL"
BASE COUNT       406 a    436 c    413 g    384 t      1 others
ORIGIN

  Query Match             82.6%;  Score 1184.6;  DB 9;  Length 1640;
  Best Local Similarity   96.9%;  Prod. No. 3.7e-257;
  Matches 1208;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 256

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||||||||||||||||||| ||||||||||||||  |||||||||
Db        257 GCAAGCCCAAATGTGTTGTGGCCAACACCTGGACACTTTCTCCTTTAGCAGTCCATCGC 316

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 180
              |||||||||||||| ||||| |||||||||||||-|||||||||||||||||| |||
Db        317 CAGTTATCACATCTATATGCAGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 376

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        377 GAACACACCTTACCTGTAAACAGAGAGACATTGAAAAGGAAGGTTAGTGGGAACCGTTGC 436

Qy        241 GCCAGCCCTGTTACTGGTCCAGTTCAAAGAGGGATGCTCACTTCTGGGCTGTCTGCAGC 300
              |||||||||||||| ||||| |||||||||||| ||||||||||||||||||||||||
Db        437 ACCAGCCCTGTTACTAGTCCAAGTTCAAAGAGGGACGCTCACTTCTGCGCTGTCTGCAGT 496

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 GATTTCGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 556

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        557 AAAACAAGCATTCAAGGCCATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 616

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 GATAAAAATCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTATGAAGTGGGA 676

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 736

Qy        541 AGTGCCGACCAGCAGCTGCACTGTGCCGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              |||||||||||||  |||||| ||| |||||||||||| |||||| ||||| |||||
Db        737 AGTGCCGACCAGCAAGTGCACTGCGCCAGCAAGGCCAAGAGGAGTGGCAGCCACACGCCC 796

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              | ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        797 CTAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTGGTGCTCACCCTCCTG 856

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              |||||||||| |||||||| |||||||||||||||| |||||||||| |||||||||
Db        857 GAGGCTGAGCCACCCCATGTACTGATCAGCCGCCCCAGCGCGCCCTTCACTGAGGCCTCC 916

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        917 ATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 976

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1036

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1037 TGTTGGATGGAGGTGTTAATGGTGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1096

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1097 CTCATCTTTGCTCCAGATCTTGTTCTGGATAGGGATGAGGGCAAATGCGTAGAAGGAATT 1156

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1157 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1216

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1217 CACAAAGAGTATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAATATGTACCCTCTG 1276

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCCTCACTTGCTGAACGCCGTG 1140
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1277 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCCCACTTGCTGAACGCCGTG 1336

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||| ||||||| |||||||| ||||||||||||||||||||||||||||||||||||
Db       1337 ACCGACGCTTTGGTCTGGGTGATCGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1396

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1397 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1443


RESULT 22
AF177936
LOCUS       AF177936                1584 bp    mRNA    linear   MAM 16 JUL 2001
DEFINITION  Ovis aries estrogen receptor beta mRNA, complete cds.
ACCESSION   AF177936
VERSION     AF177936.1  GI:5911382
KEYWORDS    .
```

```
SOURCE      Ovis aries.
  ORGANISM  Ovis aries
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 1584)
  AUTHORS   Cardenas,H.  Burke,K.A.  Bigsby,R.M.  Pope,W.F. and Nephew,K.P.
  TITLE     Estrogen receptor beta in the sheep ovary during the estrous cycle
            and early pregnancy
  JOURNAL   Biol. Reprod. 65 (1), 128-134 (2001)
  MEDLINE   21313342
   PUBMED   11420232
REFERENCE   2  (bases 1 to 1584)
  AUTHORS   Cardenas,H., Bigsby,R.M., Pope,W.F. and Nephew,K.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1999) Obstetrics and Gynecology, Indiana
            University, 1001 Walnut St., Indianapolis, IN 46202, USA
FEATURES             Location/Qualifiers
     source          1..1584
                     /organism="Ovis aries"
                     /db_xref="taxon:9940"
                     /tissue_type="ovary"
     CDS             1..1584
                     /note="steroid receptor"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAD55772.1"
                     /db_xref="GI:5911383"
                     /translation="MDVKNSPSSISNSPVSYNCGQSILPEPGPIVLPSSYVESPHFYS
                     AVTFYSPAVMNYSIPNNSEDGPGRQTTSPNVLWPTPGHLSPLATHCQSSLLYAEPQKS
                     PWCETRSLEHIFPVNPETLKPKASGSSCASPVSSPSSKFPDAHFCAVCSDYASGYHYG
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRK
                     ERGGYRIVRQRNSDFQLHCLSKTKPNGAPMTPVKEIITSALSPFQLVFTLFAFPPH
                     VLMSRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHK
                     EYLCVKAMILLNSSMYPSATASQEADSGRKLTHLLNAVTDALVWVIAKSGMSSQQQSM
                     RLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSRITGS
                     RNLAEDSESKEGSQKPQAQ"
BASE COUNT       397 a    420 c    414 g    353 t
ORIGIN

  Query Match             79.2%;  Score 1135.6;  DB 4;  Length 1584;
  Best Local Similarity   87.4%;  Pred. No. 4.4e-246;
  Matches 1243;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy          13 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT  72
               | | ||| | |  || ||||| |||| |||||||||| ||||||||||||||||||||
Db         163 AATTACAGTATTCCCAATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGCCCAAAT  222

Qy          73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT  132
               || |||||||||| ||||| ||||| |||||||||   ||||| ||||||  |||| |
Db         223 GTATTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGTCGTCACTT  282

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTA  192
               ||||||| |||||||||||||||| |||||||| | ||||  ||||||||||||||
Db         283 CTGTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAACCAGATCATTAGAACACACCTTC  342

Qy         193 CCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCACCCCTGTT  252
               ||||| |||| |||||| |||||||||| | |||||||| | |||| ||||||||||||
Db         343 CCTGTGAACAGGGAGACATTGAAAAGGAAAGCTAGTGGGAGCAGTTGTGCCAGCCCTGTT  402

Qy         253 ACTGGTCCAGGTTCAAAGAGGATGCTCACTTCTGCGCTGTCTGCAGCGATTATGCGTCG  312
               || ||||  |||||||||||||||| |||||||| ||||||||||||||||| || |||
Db         403 TCTAGTCCGAGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGCGATTATGCGTCG  462

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT  372
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         463 GGATATCACTATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT  522

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAACCGG  432
               |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||||
Db         523 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGA  582

Qy         433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGT  492
               || || ||||| |||||||| ||||| |||||||| || || || ||||||||||||||
Db         583 CGTAAAAGCTGTCAGGCCTGTCGACTGCGGAAGTGCTATGAGGTCGGAATGGTGAAGTGT  642

Qy         493 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCCGAGACAGAGAAGTGCCGACGAG  552
               |||||||||||||| || |||||||||| | |||| || |||||| | | || |||
Db         643 GGCTCCCGGACAGAAAGGTGTGGGTACCGCCATAGTGCGAAGGCACAGAAATTCTCATCAC  702

Qy         553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGGCCCGAGTGCGGAG  612
               ||||||||||      ||||  ||||||| | || | | | | |||||||  ||||
Db         703 CAGCTGCACTGCCTGAGCAAAACCAAGAGGAACGGTGCACCCATGACCCGAGTGAACGAG  762

Qy         613 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG  672
               |||||||||   |||||||||| |||||||| |||||| || ||||||||||| ||||||
Db         763 CTGCTGCTGAGTGCCCTGAGCCCCAGAGCAGCTGGTGCTTACTCTCCTGGAGGCCGAGCCG  822

Qy         673 CCCCATGTGCCTGATCAGCCGCCCCAGTCGGCCCTTCACCGAGGCCTCCATGATGATGTCC  732
               ||||| ||||| || |||||||||| |||||||||| ||||||||||||||||||||||
Db         823 CCCCACGTGCTCATGAGCCGCCCCAGCGCGCCCTTCACTGAGGCCTCCATGATGATGTCC  882

Qy         733 CTGACCAACCTGGCCGACAAGGAGCTTGTCCACATGATCAGCTGGGCCAAGAAGATTCCG  792
               |||||||| ||||||||||||||  |||||||||||||||||||||||||||||||||
Db         883 TTGACCAAGCTGCCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG  942

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGAG  852
               ||||| |||||||||||||||||||||||||||||||||||||||||||||| ||| ||||
Db         943 GGCTTCCTGCAGCTCAGCCTGTACGACCAAGTGCGGCTCTTGCAGAGCTGCTGCTTCGAG 1002

Qy         853 GTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCT  912
               ||| | ||| ||||||||||||||||||| || ||||||||||||||||||||||||||
Db        1003 GTGCTCATCGTGGGCCTGATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCT 1062

Qy         913 CCACATCTTCTTCTGGCAGAGCATCAGGCAAATCCGTAGAAGCAATTCTGCAAATCTTT  972
               ||||| || |||||||||||||||| |||||||| || |||||||||| |||||||||
Db        1063 CCAGACCTCGTTCTCGACAGGCATCAAGGCAAATGTGTTGAAGCAATTCTAGAAATCTTT 1122

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        1123 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATAT 1182

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||||| ||||| ||||||||||||||   | ||||| ||
Db        1183 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTTCAGTTACAGCATCC 1242

Qy        1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               ||||| |||||||  |||||||||| ||||| ||||||| |||||||| || ||| ||
Db        1243 CAGGAGGCTGACAGTGGCCGGAAGCTGACTCACCTGCTGAATGCCGTGACGGACGCTCTG 1302

Qy        1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAAC 1212
               || |||||||||||||||||||||||||||||||||||||||| || |||||||||||
Db        1303 GTCTGGGTGATTGCCAAGAGCGGCATGTCCTCCCAGCAACAGTCCATGCGCCTGGCTAAC 1362

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTC 1272
               || |||||||||||| || |||||||||| |||||||||||||||||||||| ||||||
Db        1363 CTGCTGATGCTCCTGTCTCATGTCAGGCATGCCAGTAACAAGGGCATGGAACACCTGCTC 1422

Qy        1273 AACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCC 1332
               |||||||||||||||||||||||| || || || || ||||||||||||||||||||||
Db        1423 AACATGAAGTGCAAAAATGTGGTGCCCGTCTACGATCTGCTGCTGGAGATGCTGAATGCC 1482

Qy        1333 CACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGAC 1392
               ||| |||||| |    |||| | |  |||| |||| ||| || || | ||| ||||||
Db        1483 CACACGCTTCGCAGCAACAAGCCTTTGGTCAAGGTCTGAGCGCAACCTGGCGGAGGAC 1542

Qy        1393 AGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
               ||| | |||||||||||||||||||| || ||| |||||||
Db        1543 AGTGAGAGCAAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1584


RESULT 23
AF110402
LOCUS       AF110402                1611 bp    mRNA    linear   MAM 11-APR-1999
DEFINITION  Bos taurus breed Holstein estrogen receptor-beta mRNA, complete
```

```
            cds.
ACCESSION   AF110402
VERSION     AF110402.1  GI:4580595
KEYWORDS    .
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1611)
  AUTHORS   Rosenfeld,C.S., Yuan,X., Manikkam,M., Calder,M.D., Garverick,H.A.
            and Lubahn,D.B.
  TITLE     Cloning, sequencing, and localization of bovine estrogen
            receptor-beta within the ovarian follicle
  JOURNAL   Biol. Reprod. 60 (3), 691-697 (1999)
  MEDLINE   99150196
   PUBMED   10026117
REFERENCE   2  (bases 1 to 1611)
  AUTHORS   Yuan,X., Rosenfeld,C.S., Manikkam,M., Calder,M.D., Garverick,H.A.
            and Lubahn,D.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Biochemistry, University of
            Missouri-Columbia, 164A Animal Sciences Research Center, 920 E.
            Campus Drive, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1..1611
                     /organism="Bos taurus"
                     /strain="breed Holstein"
                     /db_xref="taxon:9913"
                     /tissue_type="ovarian follicle"
                     /note="RT-PCR"
     CDS             28..1611
                     /function="steroid receptor"
                     /codon_start=1
                     /product="estrogen receptor-beta"
                     /protein_id="AAD24432.1"
                     /db_xref="GI:4580596"
                     /translation="MDVKNSPSSLNSPVSYNCGQSILPLEPGPTYLPSSYVESHEYS
                     AVTFYSPAVMNYSIPNNSEDGPGRQTTSPNVLWPTPGHLSPLAIHCQPSVLYAEPQKS
                     PWRETRSLEHTLPVNRETLKRKASGSSCASPATSPSSKRDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRR
                     ERCGYRIVRRQRNSDEQLHCLSKTKRNGGPMTRVKELLLSALSPEQLVLTLLEAEPPH
                     VLISRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLILDRDEGKCVEGILEIFDMLLATTSRFRELKLQHT
                     EYLCVKAMILLNSSMYPSATAPQEADSGRKLPHLLNAVTDALVWVIAKSGMSSQQQSM
                     RLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGNKSLVPGSE
                     RNLVEDSESKEGSQKPQAQ"
BASE COUNT       405 a    428 c    416 g    362 t
ORIGIN

  Query Match          78.7%;  Score 1129.2;  DB 4;  Length 1611;
  Best Local Similarity   87.1%;  Pred. No. 1.2e-244;
  Matches 1249;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy          13 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT 72
               | |  ||| | |  || |||||  |||| ||||||||||||| ||||||||||||||||||
Db         190 AATTACAGTATTCCCAATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGCCCAAAT 249

Qy          73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT 132
               || ||||||||||| |||| |||| ||||||||||   |||| |||| |||  | |  |
Db         250 GTATTGTGGCCAACTCCTGGACACCTCCTCCTTTAGCCATCCATTGCCAGCCGTCAGTT 309

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTA 192
               |||||||| |||||||||||||||| ||| |||| | ||||  ||||||||||||||||
Db         310 CTGTATGCAGAACCTCAAAAGAGTCCTTGGCGTGAAACCAGATCATTAGAACACACCTTA 369

Qy         193 CCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGCGCCAGCCCTGTT 252
               ||||| |||| |||||| |||||||||||| |||||||| | |||| ||||||||||| |
Db         370 CCTGTGAACAGGGAGACATTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGCT 429

Qy         253 ACTGGTCCAGGTTCAAAGAGGGATGCCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCG 312
               ||| |||| |||||||||||||| |||||||| |||||||||| |||| || |||
Db         430 ACTAGTCCAAGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCG 489

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 372
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db         490 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGTATT 549

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGG 432
               |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||||
Db         550 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGA 609

Qy         433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGT 492
               || || |||| |||||||| |||||||||||||| || || || ||||||||||||||||
Db         610 CGTAAAAGCTGTCAGGCCTGTCGACTTCGGAAGTGCTATGAGGTTGGAATGGTGAAGTGT 669

Qy         493 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGCAGACAGAGAAGTGCCGACGAG 552
               ||||| |||||||| || |||||||||||| | ||||| || ||||||| | | || |||
Db         670 GGCTCTCGGAGAGAAAGGTGTGGGTACCGCATAGTGCGAAGGCAGAGAAATTCTGATGAG 729

Qy         553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAG 612
               ||||||||||      |||| |||||||| | || || | |  | ||||||||  |||
Db         730 CAGCTGCACTGCCTGAGCAAAACCAAGAGGAACGGTGGACCCATGACCCGAGTGAAAGAG 789

Qy         613 CTGCTGCTGGAGCGCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 672
               |||||||||  |||||||||||| ||||||| |||| || ||||||||||| ||||||
Db         790 CTGCTGCTGAGCGCCCTGAGCCCAGAGCAGCTGGTGCTTACGCTCCTGGAGGCCGAGCCG 849

Qy         673 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 732
               ||||| |||| || ||||||||||  |||||||||| ||||||||||||||||||||||
Db         850 CCCCACGTGCTCATAAGCCGCCCCAGCAGCCCTTCACTGAGGCCTCCATGATGATGTCC 909

Qy         733 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 792
               || |||||| |||||||||||||  ||||||||||||||||||||||||||||||||||
Db         910 CTCACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG 969

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAG 852
               ||||| |||||||||||||| ||||||||||||| |||||||||| ||| |||||
Db         970 GGCTTCGTGGAGCTCAGCCTGTACGACCAAGTGCGGCTTTTGGAGAGCTGCTGGTTGGAG 1029

Qy         853 GTGTTAATGATGGGGCTGATGTGGCGCTCCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
               ||| | ||| |||||||||||||||||||| || |||||||| ||||||||||||||||||
Db        1030 GTGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCTGGCAAGCTCATCTTTGCT 1089

Qy         913 CCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTT 972
               ||||| ||  |||||||||||||| |||||||| || |||||||||| ||||||||||||
Db        1090 CCAGACCTCATTCTGGACAGGGATGAAGGGAAATGTGTTGAAGGAATTCTAGAAATCTTT 1149

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db        1150 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATAT 1209

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||||| ||||| ||||||||||||||  | |||||  ||
Db        1210 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTTCAGCTACAGCACCC 1269

Qy        1093 CAGGATGGTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               ||||| |||||||| |||||||||| ||||||| || |||| ||  || || ||  ||
Db        1270 CAGGAGGCTGACAGTGGCCGGAAGCTGACTCACCTGCTGAATGCTGTGACGGACGCTCTG 1329

Qy        1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAAC 1212
               || |||||||||||||| ||||| ||||||||||||| ||||||||||||||||||||||
Db        1330 GTCTGGGTGATTGCCAAGAGTGGCATGTCCTCCCAGCAGCAGTCCATGCGCCTGGCTAAC 1389

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTC 1272
               || |||||||||| |||||||||| || ||||||||||||||||||||||||||||||||
Db        1390 CTGCTGATGCTCCTGTCTCACGTCAGGCACGCCAGTAACAAGGGCATGGAACATCTGCTC 1449

Qy        1273 AACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCC 1332
               |||||||||||||||| |||||||| || || || ||||||||||||||||||||||||||
Db        1450 AACATGAAGTGCAAAAACGTGGTCCCGTCTACGATCTGCTGCTGGAGATGCTGAATGCC 1509

Qy        1333 CACGTGCTTCGGGGGTGCAAGTCCTCCATCGAGGGGTCCGAGTGCAGCCCGGCAGAGGAC 1392
               |||  |||||||| ||||||  |||| |||| ||| ||| || ||  ||||||
Db        1510 CACACACTTCGCGGCAACAAGTCTTTGGTCACAGGGTCTGAGCGCAACCTGGTGGAGGAC 1569
```

```
Qy       1393 AGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
             ||| | ||||||||||||||||||| || ||| ||||||||
Db       1570 AGTGAGAGCAAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1611


RESULT 24
BTA18017
LOCUS       BTA18017                1820 bp    mRNA    linear   MAM 07-SEP-1999
DEFINITION  Bos taurus mRNA for estrogen receptor beta.
ACCESSION   Y18017
VERSION     Y18017.1  GI:5823588
KEYWORDS    estrogen receptor beta.
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1820)
  AUTHORS   Walther,N., Lioutas,C., Tillmann,G. and Ivell,R.
  TITLE     Cloning of bovine estrogen receptor beta (ERbeta): expression of
            novel deleted isoforms in reproductive tissues
  JOURNAL   Mol. Cell. Endocrinol. 152 (1-2), 37-45 (1999)
  MEDLINE   99359179
   PUBMED   10432221
REFERENCE   2  (bases 1 to 1820)
  AUTHORS   Walther,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (01 SEP 1998) N. Walther, Institute for Hormone and
            Fertility, Research, Grandweg 64, D-22529 Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1..1820
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_type="granulosa cells"
     gene            154..1577
                     /gene="erbeta"
     CDS             154..1577
                     /gene="erbeta"
                     /codon_start=1
                     /product="estrogen receptor betaa"
                     /protein_id="CAB53861.1"
                     /db_xref="GI:5823589"
                     /translation "MNYSIPDNSEDGPGRQTTSPNVLWPTPGHLSPLAIHCQPSVLYA
                     EPQKSPWREIGSLAHILPVNREIINKKASGSSCASPATSPSSKRDAHFCAVCSDYASG
                     YHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVK
                     CGSRKERTGKGYRKIVRKQRNSDEGQTHCLSKTKENGGPMFKVDLELLSLASPPALVL
                     AEPPHVLISRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLE
                     SCWLEVLMVGLMWRSIDHPGKLIFAPDLILDRDEGKCVEGILEIFDMLLATTSRFREL
                     KLQHKEYLCVKAMILLNSSMYPLVTATQEADSSRKLAHLLNAVTDALVWVIAKSGMSS
                     QQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSS
                     VTGSERNLVEDSESKFGSQKPQAQ"
BASE COUNT      446 a    491 c    478 g    405 t
ORIGIN

  Query Match             78.6%;  Score 1127.6;  DB 4;  Length 1820;
  Best Local Similarity   87.1%;  Pred. No. 2.8e-244;
  Matches 1238;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy          13 ATTCCCAGCAATGTCACTAACTTGGAAGCTGGGCCTGGTCGGCAGACCACAAGCCCAAAT 72
               | |  ||| | |  |  ||||  |||| ||||||||||| ||||||||||| |||||
Db         156 AATTACAGTATTCCCGATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGTCCAAAT 215

Qy          73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT 132
               || ||||||||||| |||| |||||  ||||||||||   ||||| |||||   ||| |
Db         216 GTATTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCGGTCACTT 275

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGCTGTGGAAGCAAGATCGCTAGAACACACCTTA 192
               |||||||| |||||||||||||||| ||| |||| | |||||  ||||||||||||||
Db         276 CTGTATGCAGAACCTCAAAAGAGTCCTTGGCGTGAAACCAGATCATTAGAACACACCTTA 335

Qy         193 CCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTT 252
               ||||| |||| |||||||||||||||||| |||||||| | |||| ||||||||||| |
Db         336 CCTGTGAACAGGGAGACACTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGCT 395

Qy         253 ACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCG 312
               ||| |||| |||||||||||||||| |||||||| |||||||||| |||| || |||
Db         396 ACTAGTCCAACTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCG 455

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 372
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db         456 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGTATT 515

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCCATAAAAACCGC 432
               |||||||||||||||||||||||||||||||||||||||||||||| || |||| ||||
Db         516 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGA 575

Qy         433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGT 492
               || || |||| |||||||| |||||||||||||| || || || ||||||||||||||
Db         576 CGTAAAAGCTGTCAGGCCTGTCGACTTCGGAACTGCTATGAGGTTGGAATGGTGAAGTGT 635

Qy         493 GGCTCCCGGACAGACAGATGTGGCTACCGCCTTGTGCGGAGACAGAGAAGTGCCGAGTAG 552
               ||||| |||||||| || |||||||||||| | ||||| || ||||||| | | || |||
Db         636 GGCTCTCGGAGAGAAAGGTGTGGCTACCGCATAGTGCGAAGGCAGAGAAATTCTGATCAG 695

Qy         553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAG 612
               |||||||||||     ||||  ||||||| |  || || | |  | |||||||   |||
Db         696 CAGCTGCACTGCCTGAGCAAAACCAAGAGGAACGGTGGACCCATGACCCGAGTGAAAGAG 755

Qy         613 CTGCTGCTGGAGGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCC 672
               |||||||||  |||||||||||| |||||||| |||| || ||||||||||| |||||
Db         756 CTGCTGCTGAGCGCCCTGAGCCCAGAGCAGCTGGTGCTTACGCTCCTGGAGGCCGAGCCG 815

Qy         673 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCTCTCATGATGATGTCC 732
               ||||| ||||| || ||||||||||  |||||||||| |||||||||||||||||||||
Db         816 CCCCACGTGCTCATAAGCCGCCCCAGCACGCCTTCACTGAGGCTTCCATGATGATGTCC 875

Qy         733 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 792
               || ||||| |||||||||||  ||||||||||||||||||||||||||||||||||||
Db         876 CTCACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG 935

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAG 852
               ||||| |||||||||||||||  |||||||||||||| |||||||||| ||| |||||
Db         936 GGCTTCGTGGAGCTCAGCCTGTACGACCAAGTGCGGCTTTTGGAGAGCTGCTGGTTGGAG 995

Qy         853 GTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
               ||| | ||| ||||||||||||||||||| || |||||||| ||||||||||||||||
Db         996 GTGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCTGGCAAGCTCATCTTTGCT 1055

Qy         913 CCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTT 972
               ||||| ||  |||||||||||||||| |||||||| || |||||||||| ||||||||
Db        1056 CCAGACCTCATTCTGGACAGGGATGAAGGCAAATGTGTTGAAGGAATTCTAGAAATCTTT 1115

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db        1116 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATAT 1175

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||| |||| |||||||||||||   | |||| ||
Db        1176 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTCCAGCTACAGCACCC 1235

Qy        1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               |||||  |||||||| |||||||| |||| |||||||| || |||| || ||| ||
Db        1236 CAGGAGGCTGACAGTGGCCGGAAGCTGACTCACCTGCTGAATGCTGTGACGGACGCTCTG 1295

Qy        1153 GTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAAC 1212
               || |||||||||||||| |||| ||||||||||||| |||||||||||||||||||||
Db        1296 GTCTGGGTGATTGCCAAGAGTGGCATGTCCTCCCAGCAGCAGTCCATGCGCCTGGCTAAC 1355

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGGAGTAACAAGGGCATGGAACATCTGCTC 1272
               || |||||||||| |||||||||| || |||||||||||||||||||||||||||||||
Db        1356 CTGCTGATGCTCCTGTCTCACGTCAGGCACGCCAGTAACAAGGGCATGGAACACCTGCTC 1415

Qy        1273 AACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCC 1332
               |||||||||||||||| ||||||| || || || |||||||||||||||||||||||||
Db        1416 AACATGAAGTGCAAAAACGTGGTCCCCGTCTACGATCTGCTGCTGGAGATGCTGAATGCC 1475
```

```
Qy       1333 CACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGAC 1392
              ||| |||||||| |||||| | |||| ||||| ||| ||| || || |||||
Db       1476 CACACACTTCGCGGCAACAAGTCTTTGGTCACAGGGTCTGAGCGCAACCTGGTGGAGGAC 1535

Qy       1393 AGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||| | |||||||||||||||||||| || ||| ||||||||
Db       1536 AGTGAGAGCAAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1577


RESULT 25
AF493816
LOCUS       AF493816               1566 bp    mRNA    linear   PRI 05-JUL-2001
DEFINITION  Callithrix jacchus estrogen receptor beta 2 (ERBeta2) mRNA,
            complete cds.
ACCESSION   AF493816
VERSION     AF493816.1  GI:14600268
KEYWORDS    .
SOURCE      Callithrix jacchus.
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
            Callithrix.
REFERENCE   1  (bases 1 to 1566)
  AUTHORS   Scobie,G.A., Wilson,J.A., Millar,M.R., Macpherson,S. and
            Saunders,P.T.
  TITLE     The estrogen receptor beta variant ERBeta cx/ERBeta2 is expressed
            in a wide range of tissues in both Old and New World primates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1566)
  AUTHORS   Scobie,G.A. and Saunders,P.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) Medical Research Council, Human
            Reproductive Sciences Unit, 37 Chalmers Street, Edinburgh EH3 9EL,
            UK
FEATURES             Location/Qualifiers
     source          1..1566
                     /organism="Callithrix jacchus"
                     /db_xref="taxon:9483"
     gene            1..1566
                     /gene="ERBeta2"
     CDS             52..1512
                     /gene="ERBeta2"
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta 2"
                     /protein_id="AAK71318.1"
                     /db_xref="GI:14600269"
                     /translation="MDIKNSPSSLNSPSSYNFGQSTITPLFHGPIYIPSSYVESIHEYP
                     AMTFYSPAVMNYSIPSSVTNIFEFGPGRQIPSPNMIWSTPCHISPLAVRHQLSHEYAFP
                     QKSPWCEARSLEHTLPVSRETLKRKVSGNHCASPVTGPSSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRRERCGYRLVRRQGNAEEQLHCAGKAKRSGGHVPRVRELLLSALSPEQLVLTLLEAE
                     PPHVLISRPSVPFTEASMMMSLTKLADEELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WLEVLMVGLMWRSIDHPGKLIFAPNLLLDRDEGKCVEGILEVFDMLLATTSRFRELKL
                     QHKEYLCVKAMVLLNSQVYDPLVTATQDAESSQKLAHLLNAVTDALVWVIAKSGFSSQQ
                     QSVRLANLLMLLSHIRHASEKASHSLTSFGMEDGD"
BASE COUNT       378 a    410 c    411 g    366 t       1 others
ORIGIN

  Query Match             78.5%;  Score 1125.4;  DB 9;  Length 1566;
  Best Local Similarity   94.9%;  Pred. No. 8.9e-244;
  Matches 1171;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              |||||||||||||||||||| ||| ||||||||||||  ||||||||||||||| |
Db        211 ATGAATTACAGCATTCCCAGCAGTGTTACTAACTTGGAAGAGGGGCCTGGTCGGCAGATC 270

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||| || ||||| |||||||||| ||||||||||||||  ||||||| |
Db        271 ACAAGCCCAAATATGCTGTGGTCAACACCTGGCCACCTTTCTCCTTTAGCTGTCCATCAC 330

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 180
              |||||||||||||||||| |||||||||||||||||||||||||||||||| |||

Db        331 CAGTTATCACATCTGTATGCAGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 390

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||| |||||||||||||| ||||||||||||||||||||| |||||||||||| |||
Db        391 GAACATACCTTACCTGTAAGCAGAGAGACACTGAAAAGGAAAGTTAGTGGGAACCATTGT 450

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGGCTGTCTGCAGC 300
              |||||||||||||||||||| |||||||||||||||||||||||||| || |||||||||
Db        451 GCCAGCCCTGTTACTGGTCCAAGTTCAAAGAGGGATGCTCACTTCTGTGGCAGTCTGCAGC 510

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 570

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 630

Qy        421 GATAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        631 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGCTACGAAGTGGGA 690

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        691 ATGGTAAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGGGA 750

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              | ||||| |||||||||| || |||||||||||||||||||||||||| || |||| ||||
Db        751 AATGCCGAGGAGCAGCTGCATTGCGCCGGCAAGGCCAAGAGAAGTGGTGGTCACGTGCCC 810

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              |||||||||||||||||||  |||||||||||||||||||| ||||||||||||||||
Db        811 CGAGTGCGGGAGCTGCTGCTGAGCGCCCTGAGCCCCGAGCAGCTGGTGCTCACCCTCCTG 870

Qy        661 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||| || ||||||||||||||||||||||||||||| | ||||||||| |||||||||
Db        871 GAGGCCGAACCGCCCCATGTGCTGATCAGCCGCCCCAGCGTGCCCTTCACTGAGGCCTCC 930

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||| ||||||| |||||||||||||||||||||||||||||
Db        931 ATGATGATGTCCCTGACCAAGCTGGCCGACGAGGAGTTGGTACACATGATCAGCTGGGCC 990

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        991 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTGCGGCTCTTGGAGAGC 1050

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || ||| |||||||||||||| |||||||||||||||||||||||||||||||||||||
Db       1051 TGCTGGTTGGAGGTGTTAATGGTGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1110

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              |||||||||||||| |||| |||||||||||||||||||||||||||||||||||||||
Db       1111 CTCATCTTTGCTCCAAATCTCATTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1170

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1171 CTGGAAGTCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTGAAACTCCAA 1230

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||| |||||||||||||||||||||||| |||||||||||||       |||||||
Db       1231 CACAAAGAGTATCTCTGTGTCAAGGCCATGGTCCTGCTCAATTCCCAGTATGACCCTCTG 1290

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||| |||||||||||||| ||||||| |||||||||||||||||||||||||||||
Db       1291 GTCACAGCAACCCAGGATGCTGAGAGCAGCCAGAAGCTGGCCCACTTGCTGAATGCCGTG 1350

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || || |||||||| |||||||| || |||| ||| |||||||||||||| ||  ||
Db       1351 ACTGACGCTTTGGTCTGGGTGATCGCTAAGAGTGGCTTCTCCTCCCAGCAGCAGTCGGTG 1410

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1411 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACATCAGGCATGCGAG 1457
```

```
RESULT 26
AF267736
LOCUS       AF267736               1581 bp    mRNA    linear   MAM 15-MAR-2002
DEFINITION  Sus scrofa estrogen receptor beta (ER beta) mRNA, complete cds.
ACCESSION   AF267736
VERSION     AF267736.1  GI:14183420
KEYWORDS    .
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1581)
  AUTHORS   LaVoie,H.A., DeSimone,D.C., Gillio-Meina,C. and Hui,Y.Y.
  TITLE     Cloning and characterization of porcine ovarian estrogen receptor
            beta isoforms
  JOURNAL   Biol. Reprod. 66 (3), 616 623 (2002)
  MEDLINE   21858315
   PUBMED   11870066
REFERENCE   2  (bases 1 to 1581)
  AUTHORS   LaVoie,H.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14 MAY 2000) Cell Biology and Neurosc Iences, University
            of South Carolina, School of Medicine, Bldg 1, Columbia, SC 29208,
            USA
FEATURES             Location/Qualifiers
     source          1. .1581
                     /organism="Sus scrofa"
                     /db_xref "taxon:9823"
                     /tissue_type="ovary"
     gene            1. .1581
                     /gene="ER beta"
     CDS             1. .1581
                     /gene="ER beta"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAK15151.1"
                     /db_xref="GI:13183321"
                     /translation="MDIKNSPSNLNSPVSYNCSQSVLPLEPGFIYTPSSYVESCHEYS
                     AMTFYSPAVVNYSISSNSEVGPGRQATSPNVLWPTPGHLSPLATHCQPSLLYAEPQKS
                     PWCDTRSLEHTLPVNRFTLKRKASGSSCASPVTSPSSKRPDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRR
                     ERGGYRIVRKQRNSFGHIHCLSPAKKNGDHTTPVKELLLSTLSPPQLVLTLLEAEPPH
                     VLVSRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHE
                     EYLCVKAMILLNSSMYPSAAAQEAESSRKLTHLLNAVIDALVWVIAKSGISSQQQSVR
                     LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHILRGNKSLVPGSER
                     SRMEESESKEGSQKPQAQ"
BASE COUNT      394 a    425 c    412 g    350 t
ORIGIN

  Query Match             77.2%;  Score 1106.8;  DB 4;  Length 1581;
  Best Local Similarity   86.6%;  Pred. No. 1.4e-239;
  Matches 1232;  Conservative    0;  Mismatches  187;  Indels    3;  Gaps    1;

Qy          13 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT  72
               | ||  ||| | |  || |||||  |||| ||||||||||| ||| ||||||||| |||
Db         163 AATTACAGTATTTCCAGTAACTCCGAAGTTGGGCCTGGTCGACAGGCCACAAGCCCCAAT  222

Qy          73 GTGTTGTGGCCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT  132
               |||||||||||||| |||| |||| ||||||||||   |||| |||||   |||| |
Db         223 GTGTTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCACTT  282

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTCAAGCAAGATCGCTAGAACACACCTTA  192
               ||||||||  ||||||||||||||||||  ||||||||   |||||| ||||||||||||||
Db         283 CTGTATGCAGAACCTCAAAAGAGTCCTTGGTGTCGACACAAGATCACTAGAACACACCTTA  342

Qy         193 CCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTT  252
               |||||  ||||||||||| ||||||||||| |||||||  | |||| ||||||||||||
Db         343 CCTGTGAACAGAGAGACATTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGTT  402

Qy         253 ACTGGTCCACGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCG  312
```

```
               ||  |||||  |||||||||||||||| ||||||||  |||||||||| ||||| || ||
Db         403 ACCAGTCCAAGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCA  462

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAACAAGCATT  372
               |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         463 GGATATCACTATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAACAAGCATT  522

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACGGG  432
               |||||||||||||||||||||||||||||||||||||||||||| ||||||| || |||||
Db         523 CAAGGACATAATGATTATATTTGTCCAGCTACAAACCAGTGTACGATAGATAAGAACGGG  582

Qy         433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGAAAGTGCTATGAAGTCGGAATGGTGAAGTGT  492
               || || ||||||||||||||||||||||| ||||| || ||||| ||||||||||||||||
Db         583 CGTAAAAGCTGCCAGGCCTGCCGACTTCGAAAGTGCTATGAAGTCGGAATGGTGAAGTGT  642

Qy         493 GGCTCCCGGAGAGAGAGATGTGGCTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAG  552
               |||||||||||||| | |||||||||||| | ||||| |  |||||| | | || | |
Db         643 GGCTCCCGGAGAGAACGGTGTGGGTACCGCATAGTGCGAAAGCAGAGAAATTCTGAGGGG  702

Qy         553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAG  612
               || ||||||||     ||| |||||| | ||| |  ||| |  |||||||  ||||
Db         703 CATCTGCACTGCCTGAGCAGAGCCAAGAAGAATGGTGATCACACAACCGAGTGAAGGAG  762

Qy         613 CTGCTGGCTGCAAGGCCTGAGCCGCAGCAGCTAGTGCTCACGCTGCCTGGAGGCCGAGCCA  672
               |||||  |||  | ||||||||||  |||||||  |||||||| ||||||||||  |||||
Db         763 CTGCTCCTGAGCACCCTGAGCCCAGAGCAGCTGGTGCTCACGCTGCCTGGAGGCCGAGCCA  822

Qy         673 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC  732
               |||||  ||||||  | ||||||||||  ||||||||||||||||||||||||||||||
Db         823 CCCCACCTGCTGGTGAGCCGCCCCAGCACGCCCTTCACCGAGGCCTCCATGATGATGTCC  882

Qy         733 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG  792
               |||||||||  |||||||||||||  ||||||||||||||||||||||||||  |||||
Db         883 CTGACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAAATTCCG  942

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCCTGGAGAGCTGCTGGATGGAG  852
               ||||| ||||||||||||||| ||||||||||||| ||||||||| ||| |||||
Db         943 GGCTTCGTGGAGCTCAGCCTGATCGACCAAGTGCGGCTCCTGGAGAGCTGCTGGTTGGAG 1002

Qy         853 GTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAGCTCATCTTTGCT  912
               ||| | ||| ||||||||||||||||||| || ||||||||||||||||||||||||||
Db        1003 GTGCTCATGGTGGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCT 1062

Qy         913 CCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGAAATCTTT  972
               ||||| || ||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db        1063 CCAGACCTCGTTCTGGACAGGGATGAAGGCAAATGTGTAGAAGGAATTCTGGAAATCTTT 1122

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||| |||||||||||| ||||||||||||||||||||| || |||
Db        1123 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAGGAGTAT 1182

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCGTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||||| |||||||||||| ||| |  |||  ||
Db        1183 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTATCGTTTCAGCTGCAG---CC 1239

Qy        1093 CAGGATGGTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               || || |||| |||||||||||||| ||||  ||||||| |||||||| |||||  ||
Db        1240 CAAGAGGCTGAGAGCAGCCGGAAGCTGACTCATCTGCTGAATGCCGTGACTGATGCCCTG 1299

Qy        1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGGGCCTGGCTAAC 1212
               || |||||||  ||| || ||||||||||||||||||  ||  | |||||  |||
Db        1300 GTCTGGGTGATCGCCAGGAGTGGTATCTCCTCCCAGCAGCAGTCTGTCCAGCCTGGCCAAC 1359

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTC 1272
               ||||| |||||||| || || |||||  |||| |||||||||||||||||||  |||||
Db        1360 CTCCTCATGCTCCTTTCTCATGTCAGACATGCCAGTAACAAGGGCATGGAACACCTGCTC 1419

Qy        1273 AACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCC 1332
               |||||||||||||||||||||||||||||| || || ||||||||||||||||||||||||
Db        1420 AACATGAAGTGCAAAAATGTGGTCCCCGTCTACGACCTGCTGCTGGAGATGCTGAATGCC 1479

Qy        1333 CACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGAC 1392
               |||   ||||| ||   |||||||   |||||||||| ||| |||||| |   |||||
```

```
Db        1480 CACACTCTTCGTGGCAACAAGTCCTTGGTCACGGGGTCTGAGCGCAGCCGCATGGAGGAG 1539

Qy        1393 AGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
               ||| | ||||||||||||||||||||| || ||| |||||||
Db        1540 AGTGAGAGCAAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1581


RESULT 27
AF164957
LOCUS       AF164957                1581 bp    mRNA    linear   MAM 19-APR-2002
DEFINITION  Sus scrofa estrogen receptor beta mRNA, complete cds.
ACCESSION   AF164957
VERSION     AF164957.2  GI:14717426
KEYWORDS    .
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G., Vale-Cruz,D.S., Choi,I.,
            Katzenellenbogen,B.S., Simmen,F.A. and Simmen,R.C.
  TITLE     Molecular cloning of porcine estrogen receptor-beta complementary
            DNAs and developmental expression in periimplantation embryos
  JOURNAL   Biol. Reprod. 66 (3), 760-769 (2002)
  MEDLINE   21858833
   PUBMED   11870084
REFERENCE   2  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G. and Simmen,R.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1999) Animal Science, University of Florida,
            Shealy Drive, Building 459, Gainesville, FL 32611, USA
REFERENCE   3  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G. and Simmen,R.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2001) Animal Science, University of Florida,
            Shealy Drive, Building 459, Gainesville, FL 32611, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 13, 2001 this sequence version replaced gi:5566477.
FEATURES             Location/Qualifiers
     source          1..1581
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /dev_stage="embryo"
     CDS             1..1581
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAD45381.2"
                     /db_xref="GI:14717427"
                     /translation="MDIKNSPSNELNSPVSYNCSQSVDLPLEPGFLYIPSSYVESCHRYS
                     AMPFYSPAVVNYSISSNSEVGPGRQATSPNVIWPTPCHISPIAIRQPSLIYAEPQRS
                     PWCDTRSLEHTLPVNPETLKPKASGSSCASPVTSPSSKPDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGSRR
                     ERGGYRQIVRKQRNSEGHLHCLSKAKANGDHTTRVKELLLSTLSPEQLVTTLLEAEPPH
                     VLVSRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFMELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHK
                     EYLCVKAMILLNSSMYPSAAALESPERSRKLTHLLNAVTDALVWVIAKSGISSQQQSVR
                     LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSEE
                     SRMEESESKEGSQKPQAU"
BASE COUNT       396 a    424 c    411 g    350 t
ORIGIN

  Query Match             77.1%;  Score 1105.2;  DB 4;  Length 1581;
  Best Local Similarity   86.6%;  Pred. No. 3.3e-239;
  Matches 1231;  Conservative    0;  Mismatches  188;  Indels    3;  Gaps    1;

Qy          13 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT 72
               | |  ||| | |  || ||||  |||| |||||||||| ||| ||||||||| |||
Db         163 AATTACAGTATTTCCAGTAACTCCGAAGTTGGGCCTGGTGGACAGGCCACAAGCCCCAAT 222

Qy          73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT 132
               |||||||||||||| |||| |||| ||||||||||   |||| ||||   |||| |
Db         223 GTGTTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCACTT 282

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTA 192
               ||||||| |||||||||||||||| |||||||  |||||| ||||||||||||||||
Db         283 CTGTATGCAGAACCTCAAAAGAGTCCTTGGTGTGACACAAGATCACTAGAACACACCTTA 342

Qy         193 CCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTT 252
               ||||| |||||||||||| |||||||||||| |||||||| | |||| ||||||||||||
Db         343 CCTGTGAACAGAGACACATTGAAAAGGAAGGCTAGTGCGAGCAGTTGTGCCAGCCCTGTT 402

Qy         253 ACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCG 312
               ||  ||||| |||||||||||||||| |||||||| |||||||||| ||||| || ||
Db         403 ACCAGTCCAAGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCA 462

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 372
               |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         463 GGATATCACTATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 522

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGG 432
               |||||||||||||||||||||||||||||||||||||| ||||||| || ||||| |||||
Db         523 CAAGGACATAATGATTATATTTGTCCAGCTACAAACCAGTGTACGATAGATAAGAACCGG 582

Qy         433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGT 492
               || || ||||||||||||||||||||||| ||||| || ||||| ||||||||||||||
Db         583 CGTAAAAGCTGCCAGGCCTGCCGACTTCGAAAGTGCTATGAAGTCGGAATGGTGAAGTGT 642

Qy         493 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAG 552
               |||||||||||||  | |||||||||||| | |||| |  |||||| | || | |||
Db         643 GGCTCCCGGAGAGAACGGTGTGGGTACCGCATAGTGCGAAAGCAGAGAAATTCTGAGGGG 702

Qy         553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGCGCCCCGAGTGCGGAG 612
               || |||||||      ||| |||||||  | ||| |  ||| |  |||||||  ||||
Db         703 CATCTGCACTGCCTGAGCAGAGCCAAGAAGAATGGTGATCACACAACCGAGTGAAGCAG 762

Qy         613 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 672
               ||||| ||  | |||||||||| |||||||||||||||||||||||||||||||||||||
Db         763 CTGCTCCTGAGCACCCTGAGCCCAGAGCAGCTGGTGCTCACGCTCCTGGAGGCCGAGCCA 822

Qy         673 CCCCATGTGCTGATCAGCCGCCCCAGTGCCCCCTTCACCGAGGCCTCCATGATGATGTCC 732
               ||||| |||||| | ||||||||||  ||||||||||||||||||||||||||||||||
Db         823 CCCCACGTGCTGGTGAGCCGCCCCAGCACGCCTTCACCGAGGCCTCCATGATGATGTCC 882

Qy         733 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 792
               ||||||||| ||||||||||||  |||||||||||||||||||||||||||| |||||
Db         883 CTGACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAAATTCCG 942

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAG 852
               |||||  |||||||||||||| |||||||||||||| |||||||| ||| |||||
Db         943 GGCTTCATGGAGCTCAGCCTGTACGACCAAGTGCGGCTCCTGGAGAGCTGCTGGTTGGAG 1002

Qy         853 GTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
               ||| | ||| |||||||||||||||||| || ||||||||||||||||||||||||||||
Db        1003 GTGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCT 1062

Qy         913 CCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTT 972
               ||||| || ||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db        1063 CCAGACCTCGTTCTGGACAGGGATGAAGGGAAATGTGTAGAAGGAATTCTGGAAATCTTT 1122

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGATTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        1123 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAGGAGTAT 1182

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||||| ||||| |||||||||| |||   |   |||  ||
Db        1183 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTATCCTTCAGTTGCAG  CC 1239

Qy        1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               || || ||||| |||||||||||||| ||||  |||||| |||||||| |||| |||
Db        1240 CAAGAGGCTGAGAGCAGCCGGAAGCTGACTCATCTGCTGAATGCCGTGACTGATGCCTTG 1299

Qy        1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGGGCCTGGCTAAC 1212
               || |||||| |||| ||| || |||||||||||||||| || | |||||||| |||
Db        1300 GTCTGGGTGATCGCCAGGAGTGGTATCTCCTCCCAGCAGCAGTCTGTCCGCCTGGCCAAC 1359

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTC 1272
```

```
                |||||  |||||||| || || |||||  |||||  |||||||||||||| ||||||  |||||
Db       1360 CTCCTCATGCTCCTTTCTCATGTCAGACATGCCAGTAACAAGGGCACGGAACACTTGCTC 1419

Qy       1273 AACATGAAGTGCAAAAATGTCGTCCCAGTGTATGACCTGCTGCTGGAGATGCTCGAATGCC 1332
              |||||||||||||||||||||||||| || || ||||||||||||||||||||||||||||
Db       1420 AACATGAAGTGCAAAAATGTCGTCCCCGTCTACGACCTGCTGCTGGAGATGCTCAATGCC 1479

Qy       1333 CACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGCCAGAGGAC 1392
              |||  |||||  ||  ||||||||  |||||||||| ||| |||||  |  ||||  |||||
Db       1480 CACACTCTTCGTGGCAACAAGTCCTTGGTCACGGGGTCTGAGCGCAGCCGGCATGCAGGAC 1539

Qy       1393 AGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||| | ||||||||||||||||||||||| || ||| ||||||||
Db       1540 AGTGAGAGCAAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1581


RESULT 28
AF042058
LOCUS       AF042058                1621 bp    mRNA    linear   ROD 22-JAN-1998
DEFINITION  Rattus norvegicus estrogen receptor beta1 (ERbeta) mRNA, complete
            cds.
ACCESSION   AF042058
VERSION     AF042058.1  GI:2801690
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1621)
  AUTHORS   Petersen,D.N., Tkalcevic,G.T., Koza-Taylor,P.H., Turi,T.G. and
            Brown,T.A.
  TITLE     Identification of estrogen receptor beta2, a functional variant of
            estrogen receptor beta expressed in normal rat tissues
  JOURNAL   Endocrinology (1998) In press
REFERENCE   2  (bases 1 to 1621)
  AUTHORS   Brown,T.A. and Petersen,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1998) Molecular Sciences, Pfizer Central
            Research, Eastern Point Road, Groton, CT 06340, USA
FEATURES             Location/Qualifiers
     source          1..1621
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1..1621
                     /gene="ERbeta"
     CDS             1..1458
                     /gene="ERbeta"
                     /note="steroid hormone receptor"
                     /codon_start=1
                     /product="estrogen receptor beta1"
                     /protein_id="AAB97424.1"
                     /db_xref="GI:2801691"
                     /translation="MTFYSPAVMNYSVPGSTSNLDGGPVRQSTSPNVLWPISGHLSPI
                     ATHCQSSILYAEPQKSPWCEARSLEHTLPVNPETIKRKTSGSSCASPVTSVNAFEPAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LRKCYEVGMVKGGSRKDRRGGYRIVRPQRSSSFQVHCLSKAKRNGGHAPFVKLFTSSI
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLTHLLNAVTDAL
                     VWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHTLRGYKSSISGSESSSTEDSKNKESSQNLQSQ"
     variation       80
                     /gene="ERbeta"
                     /note="encodes leucine residue in GenBank Accession Number
                     U57439"
                     /replace "t"
     variation       813
                     /gene="ERbeta"
                     /note="encodes proline residue in GenBank Accession Number
                     U57439"
                     /replace "c"
```

```
BASE COUNT      404 a    417 c    449 g    351 t
ORIGIN

  Query Match           76.7%;  Score 1099.6;  DB 10;  Length 1621;
  Best Local Similarity   85.4%;  Pred. No. 6e-238;
  Matches 1225;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  ||||| |||    || |||| |||| |||||||||  || |||| |
Db         25 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACAGAGC 84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACTTTTCTCGTTTAGTGGTCCATCGC 120
              |||||||||||||||| | ||||||||  |||||||||  |||||||||  | ||||  ||
Db         85 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| |  || ||||| ||||||||||||||||| |||||||||||||||| ||| |
Db        145 CAATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||||||||| ||||| |||||||| | ||||
Db        205 GAGCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGCAGCAGTTGT 264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||||| |||||    |||||||||||||||||||||||| |||||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCGCCGTCTGCAGC 324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || |||| || || || ||||| ||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTT 384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||||||||||| ||||||| || ||||||||||  ||
Db        385 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 444

Qy        421 GATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || ||||||| || ||||||||||||||||||||| ||||||| |||| |||
Db        445 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 504

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| ||||||| || |||||||  | ||||||||||  | ||||||| ||||
Db        505 ATGGTCAAGTGTGGAATCCAGGAGACAACGGTGTGGGTACCTATATGTGCGGAGGCAGAGA 564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCC 600
              ||| |   ||||||| | |||||     |||| ||||||||||  || || || || |||
Db        565 AGTTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGACAAACGGTGGGCATGCACCC 624

Qy        601 CGACTGCCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| ||||||  | || |||| || |||| || ||||||||||||||||
Db        625 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 684

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCAGTGCCCCTTCACCAGCCTCC 720
              || |||| || ||| |||||||| | |||| |||||   ||||||||||||||||||||
Db        685 GAAGCTGAACCACCAATGTGCTGCTCAGCCGTCCCAGCATGCCTTCACCGAGGCCTCC 744

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||  |||| |||||||  |||||||
Db        745 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||| ||||||| |||||||| || |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGC 864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| |||| |||| ||||||||||||| || ||||||||||||||||
Db        865 TGCTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||| || ||||| || || ||||||||||||||||||||||| ||||||||||| |||
Db        925 CTCATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTGGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||| || || |||||||| || ||||||||||||||
```

```
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1044

Qy       1021 CACAAAGAATATCTCTCTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||| || ||||||||||| ||||||||||||||| ||||| ||||||||||||||  ||
Db       1045 CACAAGCAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTG 1104

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |   | || | |||||| || || || ||||||||||| | ||| | |||||||| |||
Db       1105 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1164

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||||| ||||| || ||||||||||||||||||  ||  |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || ||||| |||||||||||||| || || |||||||||||    ||||||||||||||||
Db       1225 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATG 1284

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              |||||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1285 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAG 1344

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              |||||||||| ||| ||||||| |||| |||||||| ||| ||||||| |||||||||
Db       1345 ATGCTGAATGCTCACAGCCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGC 1404

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              |  |||||||||||| | |||||| |||||||||||| ||||||||||||
Db       1405 TCAACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGA 1458


RESULT 29
RNAJ2602
LOCUS       RNAJ2602                1650 bp    DNA     linear   ROD 18-APR-1998
DEFINITION  Rattus norvegicus mRNA for estrogen receptor beta (clone
            pcDNA.ERbeta).
ACCESSION   AJ002602
VERSION     AJ002602.1  GI:3077648
KEYWORDS    estrogen receptor beta.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) Aldridge T.C., Molecular Endocrinology
            Group, ZENECA Central Toxicology Laboratory, Alderley Park,
            Macclesfield, Cheshire SK10 4TJ, ENGLAND
  REMARK    revised by [3]
REFERENCE   2  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Tissue-specific responses to estrogen: an explanation based on
            differential activation of multiple estrogen receptors with
            different estrogen response elements
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1997) Aldridge T.C., Molecular Endocrinology
            Group, ZENECA Central Toxicology Laboratory, Alderley Park,
            Macclesfield, Cheshire SK10 4TJ, ENGLAND
COMMENT     On Apr 24, 1998 this sequence version replaced gi:2632166.
FEATURES             Location/Qualifiers
     source          1..1650
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone="pcDNA.ERbeta"
                     /tissue_type="Prostate"
     gene            1..1650
                     /gene="ERbeta"
     CDS             1..1650
                     /gene="ERbeta"
                     /note="The clone was isolated by RT-PCR from total RNA"
                     /codon_start=1
                     /evidence=experimental
                     /product="estrogen receptor beta"
                     /protein_id="CAA05631.1"
                     /db_xref="GI:3077649"
                     /db_xref="SPTREMBL:O35784"
                     /translation="MSICTSSHKFFSQLRPSFDMFLKNSPSSLSSPASYNCSQSILPL
                     EHGPIYIPSSYVDNRHEYSAMTFYSPAVMNYSVPGSTSNLDGGFVRQSTSPNVLWPTS
                     GHLSPLATHCQSSLLYAEPQKSPWCEAPSLEHTLPVNPETLKPKLSGSSCASPVTSPN
                     AKPDAHFCAVCSDYASGYHYGVWPCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRK
                     SCQACRLRKCYEVGMVKCGSPRPRCGYRTVRPQRSSSEQVHCLSKAEKNGCHAPPVKE
                     LLLSTLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKK
                     IPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGI
                     LEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLLHLLN
                     AVTDALVWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYD
                     DLLEMLNAHTLRGVKPSISCSFCSSTFDSKNKFSSQNLQSQ"
BASE COUNT      415 a    443 c    432 g    360 t
ORIGIN

  Query Match          76.6%;  Score 1098;  DB 10;  Length 1650;
  Best Local Similarity   85.4%;  Pred. No. 1.4e-237;
  Matches 1224;  Conservative    0;  Mismatches  210;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||| |||     || |||| |||| ||||||||||  || |||| |
Db        217 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACAGAGC 276

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||||| | ||||||||  |||||||||| ||||||||| | ||||| ||
Db        277 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 336

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCTGGTGTGAAGCAAGATCACTA 180
              || | ||| |  || |||| ||||||||||||||||| |||||||||||||||| |||
Db        337 CAATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 396

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||||||| |||||| |||||||| | ||||
Db        397 GAGCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGT 456

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGC 300
              |||||||||||||||| |||||    |||||||||||||||||||||||||| ||||||||
Db        457 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCGCCGTCTGCAGC 516

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || ||||| || || || ||| | ||||||||||||||||||||||||
Db        517 CATTATGCATCTGGGTATCATTACGGCGTTTGGCCATGTGAAGGATGTAAGGCCTTTTTT 576

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||||||||||||| ||||||| || ||||||||||
Db        577 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 636

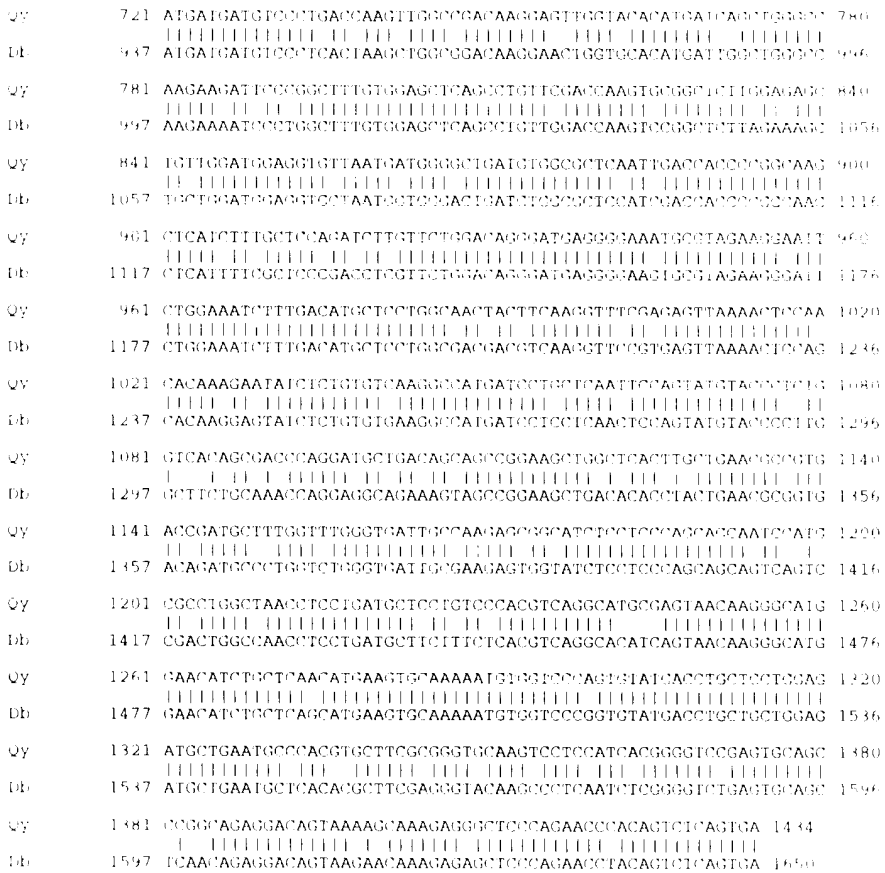
Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| || ||||||||||||||||||||| |||||||| ||||| |||
Db        637 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 696

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGACAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| |||||||  | ||||||||||  | |||||||| ||||||
Db        697 ATGGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGA 756

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||| |   ||||||| | |||||     |||| ||||||||||  || || ||||| |||
Db        757 AGTTCTACGGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCACGCACCC 816

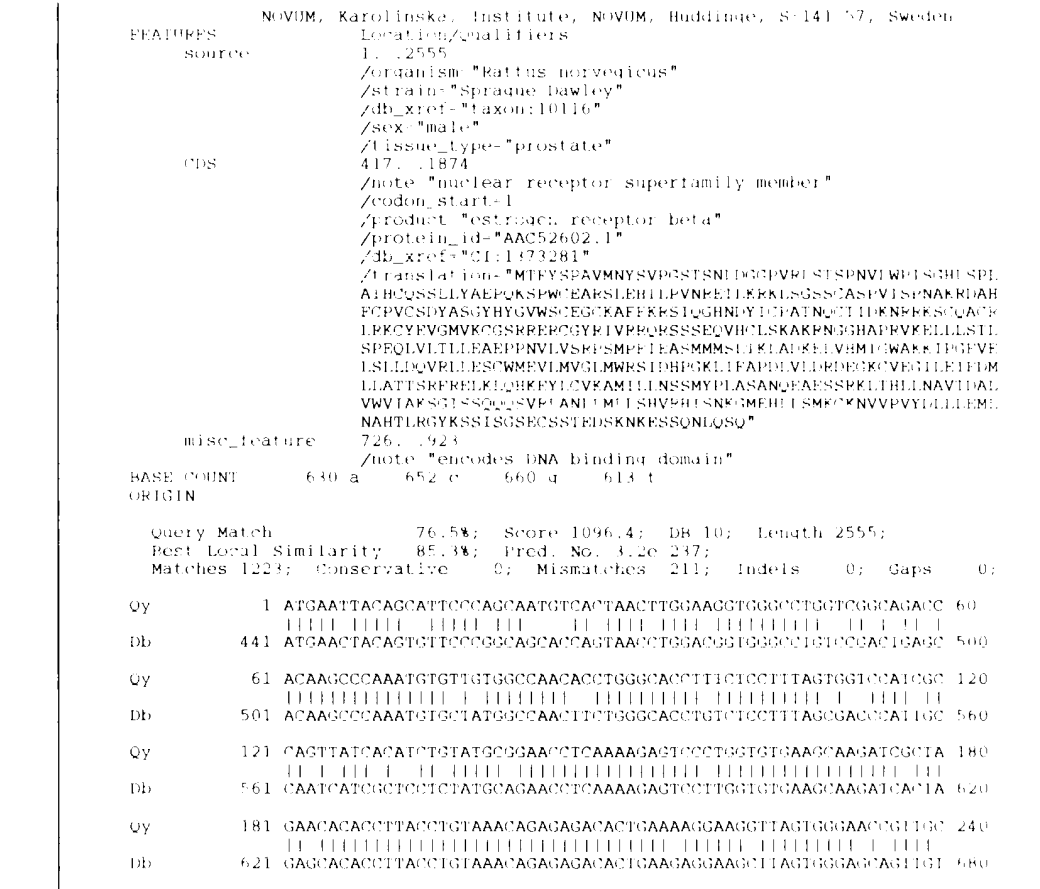
Qy        601 CGAGTGCGGGAGCTGCTCTGGACGCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| ||||||  | || |||| || ||||| || ||||||||||||||||
Db        817 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 876

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || ||||| || ||| |||||||| | ||||| |||||   ||||||||||||||||||
Db        877 GAAGCTGAACCAACCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 936
```



RESULT 30
RNU57439
LOCUS RNU57439 2555 bp mRNA linear RDN 12-JUN-1996
DEFINITION Rattus norvegicus estrogen receptor beta mRNA, complete cds.
ACCESSION U57439
VERSION U57439.1 GI:1373280
KEYWORDS .
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 8 to 2555)
AUTHORS Kuiper,G.G., Enmark,E., Pelto-Huikko,M., Nilsson,S. and Gustafsson,J.A.
TITLE Cloning of a novel receptor expressed in rat prostate and ovary
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (12), 5925-5930 (1996)
MEDLINE 96234066
PUBMED 8650195
REFERENCE 2 (bases 1 to 2555)
AUTHORS Enmark,E.L.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-1996) Eva L. Enmark, Center for Biotechnology, NOVUM, Karolinska, Institute, NOVUM, Huddinge, S-141 57, Sweden
FEATURES Location/Qualifiers
source 1..2555
/organism="Rattus norvegicus"
/strain="Sprague Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="prostate"
CDS 417..1874
/note "nuclear receptor superfamily member"
/codon_start=1
/product "estrogen receptor beta"
/protein_id="AAC52602.1"
/db_xref="GI:1373281"
/translation="MTFYSPAVMNYSVPGSTSNEIGGPVRLSISPNVIWPESGHLSPL
AIHGQSSLLYAEPGKSPWCEAKSLEHLLPVNPELLERKLSGSSCASPVTSPSNAKRDAH
FCPVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQAC
LPKCYEVGMVKGGSRKERRGYRLVRPQRSSSEQVHCLSKAKRNGGHAPRVKELLSTL
SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMISWAKKIPGFVE
LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLIHLLNAVIDAL
VWVIAKSGISSQQQSVPLANLLMLLSHVRHLSNKGMEHLLSMKCKNVVPVYDLLLEML
NAHTLRGYKSSISGSECSSTEDSKNKESSQNLQSQ"
misc_feature 726..923
/note "encodes DNA binding domain"
BASE COUNT 630 a 652 c 660 g 613 t
ORIGIN

Query Match 76.5%; Score 1096.4; DB 10; Length 2555;
Best Local Similarity 85.3%; Pred. No. 3.2e-237;
Matches 1223; Conservative 0; Mismatches 211; Indels 0; Gaps 0;



```
                 || |||  |||||| ||||||  | || |||| || ||||| || ||||||||||||||||
Db        1041 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 1100

Qy         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
               || ||||| || ||| |||||||| | |||||  |||||   ||||||||||||||||||
Db        1101 GAACGTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 1160

Qy         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
               |||||||||||||| || ||| |||| ||||||||  |||| ||||||||  ||||||||
Db        1161 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 1220

Qy         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
               ||||| || || ||||||||||||||||||||||| |||||||| |||||||| || |||
Db        1221 AAGAAAATCCCCTGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCCGGCTCTTAGAAAGC 1280

Qy         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
               || |||||||||||| |||| |||| |||||||||||||| || |||||||||||||||
Db        1281 TGCTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 1340

Qy         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
               ||||| || ||||| || || |||||||||||||||||||||||| ||||||||||| |||
Db        1341 CTCATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGGATT 1400

Qy         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
               |||||||||||||||||||||| || || |||||||| || |||||||||||||||||
Db        1401 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1460

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
               ||||| || |||||||||||| ||||||||||||| ||||| ||||||||||||||  ||
Db        1461 CACAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTG 1520

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
               |   | || | |||||| || || || |||||||||||| | ||| | |||||||| |||
Db        1521 GCTTCTGCAAACCAGGAGCCAGAAAGTAGCCGGAAGCTGACACACCTACTCAACGCGGTG 1580

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               || ||||  |||| |||||||||||| ||||| || |||||||||||||||| ||  |
Db        1581 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1640

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
               || ||||| |||||||||||||| || || |||||||||||    ||||||||||||||||
Db        1641 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATG 1700

Qy        1261 GAACATCTGCTCAACATCAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
               |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        1701 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAG 1760

Qy        1321 ATGCTGAATGCCCACCGTGCTTCGCGGGTCCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
               |||||||||||| |||  |||||| |||| |||||| | ||| |||||| |||||||||||
Db        1761 ATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGC 1820

Qy        1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
                 |  |||||||||||||| | |||||| |||||||||||| ||||||||||||
Db        1821 TCAACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGA 1874


RESULT 41
A61584
LOCUS       A61584                  2568 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 1 from Patent WO9709348.
ACCESSION   A61584
VERSION     A61584.1  GI:4715916
KEYWORDS    .
SOURCE      Rattus rattus.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2568)
  AUTHORS   Kuiper,G.G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348 A 1 14-MAR 1997;
            KAROBIO AB (SE)

FEATURES             Location/Qualifiers
     source          1..2568
                     /organism "Rattus rattus"
                     /db_xref "taxon:10117"
BASE COUNT      634 a    654 c    663 g    617 t
ORIGIN

  Query Match          76.5%;  Score 1096.4;  DB 6;  Length 2568;
  Best Local Similarity   85.3%;  Pred. No. 3.2e-247;
  Matches 1224;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGGCCTGGTCGGCAGACC 60
              ||||| ||||| ||||| ||| ||  || |||| |||| |||||||||| || | || |
Db        448 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGC 507

Qy         61 ACAACCCCAAATGTGTTGTGGCCAACACCTGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | |||||||  |||||||| |||||||| | |||| ||
Db        508 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCATTGC 567

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| |  || ||||| |||||||||||||||| ||||||||||||||| |||
Db        568 CAATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 627

Qy        181 GAACACACCTTACCTGTAAACAGAGACACTCAAAAGGAAGGTTACTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||||| |||||| |||||||| | ||||
Db        628 GAGCACACTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGT 687

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||| |||||    ||||||||||||||||||||||| | |||||||||
Db        688 GCCAGCCCTGTTACTAGTCCAAAGGCAAAGAGGGATGCTCACTTCTGCCCGTCTGCAGC 747

Qy        301 GATTATGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| |||||| || ||||| || || || |||| ||||||||||||||||||||||||
Db        748 GATTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTT 807

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||| |||||||| || ||||||||| ||
Db        808 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 867

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || ||||||| || ||||||||||||||||||||||| |||||||| |||| |||
Db        868 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 927

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| |||||||  | ||||||||||  | |||||||| ||||||
Db        928 ATGGTCAAGTGTGCATCCAGGAGAAGCGGTGTGGGTACCGTATAGTGCGGGAGGCAGAGA 987

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||| |   |||||| | ||||     |||| ||||||||||  || || || || |||
Db        988 AGTTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCC 1047

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| ||||||  | || |||| || |||| || ||||||||||||||||
Db       1048 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 1107

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || ||||| || ||| |||||||| | |||||  |||||   ||||||||||||||||||
Db       1108 GAAGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 1167

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||  |||| ||||||||  ||||||||
Db       1168 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 1227

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || ||||||||||||||||||||||| |||||||| |||||||| || |||
Db       1228 AAGAAAATCCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGC 1287

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| |||| |||| |||||||||||||| || |||||||||||||||
Db       1288 TGCTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 1347

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||| || ||||| || || |||||||||||||||||||||||| ||||||||||| |||
```

```
Db       1348 CTCATTTTCGCTCCCGACCTGGTTCTGGACAGGGATGAGGGCAAGTGCGTAGAAGGGATT 1407

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||| || || |||||||| || ||||||||||||||
Db       1408 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1467

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||| || |||||||||| ||||||||||||| ||||| ||||||||||||||  ||
Db       1468 CACAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTTG 1527

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |   | || | |||||| || || || ||||||||||| | ||| | ||||||| |||
Db       1528 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1587

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||||| ||||| || ||||||||||||||||| || |
Db       1588 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1647

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || ||||| |||||||||||| || || ||||||||||    ||||||||||||||||
Db       1648 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATG 1707

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              |||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       1708 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAG 1767

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              |||||||||| ||| ||||| |||| |||||||| ||| |||||| ||||||||||||||
Db       1768 ATGCTGAATGCTCACACGCTTCCAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGC 1827

Qy       1381 CGGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              |  |||||||||||| | ||||||| |||||||||||| ||||||||||||||
Db       1828 TCAACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGA 1881


RESULT 42
AR075920
LOCUS       AR075920               2568 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 1 from patent US 5958710.
ACCESSION   AR075920
VERSION     AR075920.1  GI:10002666
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2568)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J.-A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710-A 1 28 SEP 1999;
FEATURES             Location/Qualifiers
     source          1. .2568
                     /organism "unknown"
BASE COUNT      644 a    654 c    663 g    617 t
ORIGIN

  Query Match           76.5%;  Score 1096.4;  DB 6;  Length 2568;
  Best Local Similarity   85.4%;  Pred. No. 3.2e-237;
  Matches 1223;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  ||||| |||    || |||| |||| |||||||||  || | || |
Db        448 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGC 507

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | |||||||  |||||||||| ||||||||| |  |||| ||
Db        508 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 567

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCCCTA 180
              || | ||| |  || ||||| ||||||||||||||||| ||||||||||||||||| |||
Db        568 CAATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 627

Qy        181 GAACACACGTTACCTGTAAACAGACGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||||||| |||||| |||||||| | ||||

Db        628 GAGCACACGTTACCTGTAAACAGACGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGT 687

Qy        241 GCCAGCCCTGTTACTAGTCCAAAGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGC 300
              ||||||||||||||| |||||    |||||||||||||||||||||| | ||||||||
Db        688 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGC 747

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || ||||| || || || ||||| ||||||||||||||||||||||||
Db        748 GATTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTT 807

Qy        361 AAAACAAGCATTCAAGCACATAATCATTATATTTGTCCACCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||||||||||| |||||||| || ||||||||| ||
Db        808 AAAACAAGCATTCAAGCACATAATGATTATATGTGTCCACCCACCAATCAGTGTACCATA 867

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGCAAGTGTTACGAAGTGGGA 480
              || || |||||||| || ||||||||||||||||||||| |||||||| ||||| |||
Db        868 GACAACAACCGCGCTAAAAGCTGCCAGGCCTGCCGACTTCGCAACTGCTATGAACTACGA 927

Qy        481 ATGGTGAAGTGTGGCTCCCCGGAGAGAGAGATGTGGCTACCGCCTTCTGCCGGAGACAGAGA 540
              ||||| |||||||| ||| ||||||| | ||||||||||| | |||||||| ||||||
Db        928 ATGGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCCGAGGCAGAGA 987

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGCGCC 600
              ||| |   |||||||  | |||||     |||| ||||||||||  || || || || |||
Db        988 AGTTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCC 1047

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| ||||||  | || |||| || |||| || ||||||||||||||
Db       1048 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 1107

Qy        661 GAGGCTGAGCCGCCCCATGCTCTACAGCGAGTTCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || |||| || |||||  |||||||  |||| |||||||  ||||||||||||||||||
Db       1108 GAAGCTGAACCACCCAATGTGCTGGTGAGCCCTCCCAGCATGCCCTTCACCGAGGCCTCC 1167

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||  |||| ||||||||  ||||||||
Db       1168 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 1227

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||| |||||||| ||||||||  || |||
Db       1228 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGC 1287

Qy        841 TGTTGGATGGAGCTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||| ||||| |||| ||||||||||||| || ||||||||||||||||||
Db       1288 TGCTGGATGCAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 1347

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||| || ||||| || || ||||||||||||||||||||||| |||||||||| |||
Db       1348 CTCATTTTCGCTCCCGACCTGGTTCTGGACAGGGATGAGGGCAAGTGCGTAGAAGGGATT 1407

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||| || || |||||||| || ||||||||||||||
Db       1408 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1467

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||| || |||||||||| ||||||||||||| ||||| ||||||||||||||  ||
Db       1468 CACAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTTG 1527

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |   | || | |||||| || || || ||||||||||| | ||| | ||||||| |||
Db       1528 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1587

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||||| ||||| || ||||||||||||||||| || |
Db       1588 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1647

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || ||||| |||||||||||| || || ||||||||||    ||||||||||||||||
Db       1648 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATG 1707

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              |||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       1708 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAG 1767
```

```
Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
             ||||||||||| || |||||| |||| |||||||||| || |||||||| ||||||||||
Db       1768 ATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGC 1827

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
               | |||||||||||||| | |||||| ||||||||||||| ||||||||||||
Db       1828 TCAACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGA 1881


RESULT 33
AX022455
LOCUS       AX022455               2568 bp    DNA     linear   PAT 20-SEP-2000
DEFINITION  Sequence 1 from Patent EP0935000.
ACCESSION   AX022455
VERSION     AX022455.1  GI:10046068
KEYWORDS    .
SOURCE      black rat.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae;
            Rattus.
FEATURES             Location/Qualifiers
     source          1..2568
                     /organism="Rattus rattus"
                     /db_xref="taxon:10117"
BASE COUNT      644 a    654 c    664 g    617 t
ORIGIN

  Query Match             76.5%;  Score 1096.4;  DB 6;  Length 2568;
  Best Local Similarity   85.3%;  Pred. No. 3.2e-247;
  Matches 1223;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGGCCTGGTCGGCAGACC 60
              ||||| ||||| ||||| ||| || || |||| |||| |||||||||| || | || |
Db        448 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGGCCTGTCCGACTGAGC 507

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATGGC 120
              |||||||||||||||| | |||||||| |||||||||| |||||||||| | |||| ||
Db        508 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 567

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATGGCTA 180
              || | ||| |  || |||| |||||||||||||||| |||||||||||||||||| |||
Db        568 CAATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 627

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAGGAAGGTTAGTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||||| |||||| |||||||| | ||||
Db        628 GAGCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGT 687

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGC 300
              |||||||||||||| |||||    ||||||||||||||||||||||| | ||||||||||
Db        688 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGC 747

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGCGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || ||||| || || || ||||| ||||||||||||||||||||||||
Db        748 GATTATGCATCTGGGTATCATTACGGCGTTGGTCATGTGAAGGATGTAAGGCCTTTTTT 807

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||||||| |||||||| || |||||||||| ||
Db        808 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 867

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAACGTTACGAAGTGGGA 480
              || || |||||||| || |||||||||||||||||||| |||||||| ||||| |||
Db        868 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 927

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| ||||||| | |||||||||| | |||||||| ||||||
Db        928 ATGGTCAACTGTGGATCCAGGAGAAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGA 987

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCC 600
              ||| |   ||||||| | |||||     |||| |||||||||| || || || || |||
Db        988 AGTTCTAGCGAGCAGCTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCC 1047

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || ||| ||||| |||||| | || |||| || |||| || ||||||||||||||||
Db       1048 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 1107

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGGCTCC 720
              || ||||| || ||| |||||||| | ||||| ||||| |||||||||||||||||||
Db       1108 GAAGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGGCTCC 1167

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||| |||| |||||||| ||||||||
Db       1168 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 1227

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||| ||||||| ||||||| || |||
Db       1228 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAACTCCGGCTCTTAGAAAGC 1287

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||
Db       1288 TGCTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 1347

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||| || |||| || || ||||||||||||||||||||| |||||||||| |||
Db       1348 CTCATTTTGCCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGATT 1407

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||| || || |||||||| || ||||||||||||||
Db       1408 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1467

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||  || |||||||||| |||||||||||||| |||||  ||||||||||||||  ||
Db       1468 CACAAGGAGTATCTCTGTGTGAAGGCCATGATCCTGCTCAACTCCAGTATGTACCCTCTG 1527

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |   | || | ||||| || || || |||||||||||| | ||| | |||||||  |||
Db       1528 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1587

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||||| ||||| || |||||||||||||||||| ||  |
Db       1588 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1647

Qy       1201 CGGCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || ||||| |||||||||||||| || || |||||||||||    ||||||||||||||
Db       1648 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATG 1707

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       1708 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAG 1767

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              |||||||||||| |||  |||||| |||| |||||||||| || |||||||| ||||||||
Db       1768 ATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGC 1827

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
               | |||||||||||||| | |||||| ||||||||||||| ||||||||||||
Db       1828 TCAACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGA 1881


RESULT 34
MMU81451
LOCUS       MMU81451               2152 bp    mRNA    linear   ROD 01-APR-1997
DEFINITION  Mus musculus estrogen receptor beta (Estrb) mRNA, complete cds.
ACCESSION   U81451
VERSION     U81451.1  GI:1912467
KEYWORDS    nuclear receptor; steroid receptor; transcription factor; mouse
            chromosome 12.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2152)
  AUTHORS   Tremblay,G.B., Tremblay,A., Copeland,N.G., Gilbert,D.J.,
            Jenkins,N.A., Labrie,F. and Giguere,V.
  TITLE     Cloning, chromosomal localization, and functional analysis of the
```

```
            murine estrogen receptor beta
  JOURNAL   Mol. Endocrinol. 11 (3), 353-365 (1997)
  MEDLINE   97211483
   PUBMED   9058381
REFERENCE   2  (bases 1 to 2152)
  AUTHORS   Tremblay,G.B., Tremblay,A. and Giguere,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1996) Molecular Oncology Group, Royal Victoria
            Hospital, 687 Pine Avenue West, Montreal H3A 1A1, Canada
FEATURES             Location/Qualifiers
     source          1. .2152
                     /organism "Mus musculus"
                     /strain "129/Sv"
                     /db_xref "taxon:10090"
                     /chromosome="12"
                     /tissue_type="ovary"
     gene            1. .2152
                     /gene="Estrb"
     CDS             12. .1469
                     /gene="Estrb"
                     /function="nuclear hormone receptor transcription factor;
                     steroid receptor"
                     /note="ERbeta"
                     /codon_start 1
                     /product "estrogen receptor beta"
                     /protein_id "AAB51132.1"
                     /db_xref "GI:1912468"
                     /translation="MTFYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPISGHLSPL
                     ATHCQSSLLYAEPQKSPWCEARSLEHTLPVNRFTLKRKLGGSGGFASPVTSPSAKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LPKCYEVGMVKGGSRKERGGYRTVRRQRSASEQVHCLNKAKRTSCETPRVKFLLNSL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMICWAKKIPGFVE
                     LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEAESSRKLTHLLNAVTDAL
                     VWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEM
                     NAHTLRGYKSSISGSEDQSHIEDSKSKEGSQNLQSQ"
BASE COUNT       549 a    526 c    570 g    507 t
ORIGIN

  Query Match          75.8%;  Score 1086.8;  DB 10;  Length 2152;
  Best Local Similarity   84.9%;  Pred. No. 4.8e-235;
  Matches 1217;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||||||||    | |||| ||||||||||||||| || |||||
Db         36 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGCCAGACT 95

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||| | |||||||  |||| |||| |||||||||    ||  ||
Db         96 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 155

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| | ||| |||| ||||||||||||||||| ||||||||||||||| |||
Db        156 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 215

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTGGC 240
              |||||||||| |||||||||||||||| |||| |||||| || | |||| | ||||
Db        216 GAACACACCTTGCCTGTAAACAGAGAGACCCTGAAGAGGAAGCTTGGCGGGAGCGGTTGT 275

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGC 300
              |||||||||||||| |||| |  | |||||||||||||||||||||| |||||||||
Db        276 GCCAGCCCTGTTACTAGTCCAAGCGCCAAGAGGGATGCTCACTTCTGCGCCGTCTGCAGT 335

Qy        301 GATTACGCATCGGCATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| |||| || |||| || || |||||||| ||||||||||||||||||||||||
Db        336 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 395

Qy        361 AAAAGAAGCATTCAAGGACATAATGATATATTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||| ||||| |||||||| || |||||||||| ||
Db        396 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACCATA 455

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || ||||||| || ||||||||||||||||||||||| ||||||||||||| |||

Db        456 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 515

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGACGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| |||||| || |||||||||||  | || || ||||||||
Db        516 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA 575

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||  ||||||| |||| ||       ||| ||||||||||   | || ||| | |||
Db        576 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCAGTGGGCACACACCC 635

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| |||||| || | |||| ||||||||||| ||||||||||| |||
Db        636 CGGGTGAAGGACTTCTGTTCCAAGCCTTTGAGTGCCAACAGCTTCTGCTCACTCTCCTG 695

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || ||||||| ||| |||||||  | |||| ||||   |||||||||||||||||||
Db        696 GAAGCTGAGCCACCCAATGTGCTAGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 755

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTCCACATGATCAGCTGGGCC 780
              |||||||||||| || ||| |||| |||||||  |||| ||||||||  ||||||||
Db        756 ATGATGATGTCCCTCACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 815

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||||| ||||||| || |||||||| |||
Db        816 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAAAGC 875

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| | ||| |||||||||||||||||||| || ||||||||||||||
Db        876 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAG 935

Qy        901 CTCATCTTTGCTCCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              |||||||||||||||| || |||||||||||||||||||||| |||| |||| |||
Db        936 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGCATGAGGGGAAGTGCGTGGAAGGGATT 995

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||| ||||||||||| || ||  || |||| || |||||||||| ||
Db        996 CTGGAAATCTTTGGCATGCTCCTGCCGACGACGGCACGCTTCCGTGAGTTAAAACTGCAG 1055

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||||||||| |||| |||||||||| || |||| ||||||||||||  ||
Db       1056 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCCTCTG 1115

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |  || || | |||||| || || |||||||||||| | ||| |  |||||| |||
Db       1116 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1175

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||| | |||| || |||| |||||||||| ||  |
Db       1176 ACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTGGAATCTCTTCCCAGCAGCAGTCAGTC 1235

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || |||| |||||||||||| || || || |||||||    ||||||||||||||
Db       1236 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGGCACATCAGTAACAAGGGCATG 1295

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||| ||||||||||||||||||||||| |||| ||||||||||||||
Db       1296 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTACGACCTGCTGCTGGAG 1355

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              |||||||||| ||  |||||| |||| |||||||| ||| |||||| |||||| ||
Db       1356 ATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCTGC 1415

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              || ||||||||||||| ||||||||||||||||||||||  |||| ||||||
Db       1416 TCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAGAACCTCCAGTCACAGTGA 1469


RESULT 35
A61588
LOCUS       A61588                  1458 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 6 from Patent WO9709348.
ACCESSION   A61588
VERSION     A61588.1  GI:3715918
```

```
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Kuiper,G.G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348 A 6 13-MAR 1997;
            KABIGEN AB (SE)
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT      365 a    374 c    404 g    317 t
ORIGIN

  Query Match             75.3%;  Score 1080.4;  DB 6;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 1.3e-243;
  Matches 1213;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||||||||    1  |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGCCAGACT 84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
               |||||||||||||| | ||||||||  |||| ||||| ||||||||||    ||| ||
Db         85 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATGGCTA 180
              || | ||| | ||| |||||  ||||||||||||||||  |||||||||||||||||| |||
Db        145 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGC 240
              |||||||||||  |||||||||||||||||  |||||  ||||||  || | ||||  | ||||
Db        205 GAACACACCTTGCCTGTAAACAGAGAGACCCTGAAGAGGAAGCTTGGCGGGAGCGGTTGT 264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||||| |||||  |  | ||||||||||||||||||||| ||  ||||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGATGCTCACTTCTGTGCCGTCTGCAGT 324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| |||||  || |||||  || ||  |||||||||||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||| |||||  |||||||  || ||||||||||  ||
Db        385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA 444

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || ||  |||||||  | |  |||||||||||||||||||  ||||||||||||  |||
Db        445 GACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 504

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGGAGACAGAGA 540
              |||||  |||||||| |||  |||||||  ||  ||||||||||   |  || ||  ||||||||
Db        505 ATGGTCAACTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA 564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACGCGCCC 600
              ||||||  |||||||  |||  ||       |||  ||||||||||   |  ||  |  |||
Db        565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCAGTGGGCACACACCC 624

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||  |||  ||||||  |||||  || |  ||||  ||||||||||  ||||||||||  |||
Db        625 CGGGTGAAGGAGCTACTGCTGAACTCTCTGATCCCGAGCAGCTGGTGCTCACCCTGCTG 684

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || ||||||||  ||| ||||||  |  || ||  |||||   ||||||||||||||||||||
Db        685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| ||  ||| |||| ||||||||  |||| ||||||||  ||||||||
Db        745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              |||||  ||  ||  |||||||||||||||||||||||||  |||||||  ||  |||||| |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC 864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||  |||||||||||  | |||  |||||||||||||||||||  ||  |||||||||||||||
Db        865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||||| ||  ||||||||||||||||||||||||| |||||  |||||  |||
Db        925 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGATGAGGGCAAGTGCGTGGAAGGGATT 984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||||||||| || ||  ||  |||| ||  ||||||||  ||
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTGCAG 1044

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||||||||| |||||  |||||||||| ||  |||||  |||||||||||||   ||
Db       1045 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG 1104

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |  || || | |||||| || || ||  ||||||||||| | ||| |  |||||||  |||
Db       1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1164

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||  | |||||  | ||||  |||||||||||  || |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGAATCTCTTCCCAGCAGCAGTCAGTC 1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || |||||  |||||||||||||| || || ||  ||||||||    ||||||||||||||
Db       1225 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGGCACATCAGTAACAAGGGCATG 1284

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||  |||||||||||||||||||||||||||||  |||||  ||||||||||||
Db       1285 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTACGACCTGCTGCTGGAG 1344

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||| |||  ||||||  |||| ||||||||| ||| |||||||  | |||| ||
Db       1345 ATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGGGTGCTGC 1404

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
               || |||||||||||| ||||||||||||||||||||||  ||||||||||||
Db       1405 TCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAGAACCTCCAGTCTCAGTGA 1458


RESULT 46
AR075922
LOCUS       AR075922               1458 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 6 from patent US 5958710.
ACCESSION   AR075922
VERSION     AR075922.1  GI:10002668
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J. A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710-A 6 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="unknown"
BASE COUNT      365 a    374 c    404 g    317 t
ORIGIN

  Query Match             75.3%;  Score 1080.4;  DB 6;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 1.4e-243;
  Matches 1213;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||||||||    1  |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGCCAGACT 84
```

```
Qy        61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
             ||||||||||||||| | ||||||||  |||| ||||| ||||||||||     ||| ||
Db        85 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 144

Qy       121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
             || | ||| | ||| ||||| |||||||||||||||| ||||||||||||||||||| |||
Db       145 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy       181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
             |||||||||||| |||||||||||||||||| ||||| ||||| || | |||| | ||||
Db       205 GAACACACCTTGCCTGTAAACAGAGAGACCCTGAAGAGGAAGCTTGGCGGGAGCGGTTGT 264

Qy       241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGC 300
             |||||||||||||| ||||| |  | ||||||||||||||||||||||| || ||||||||
Db       265 GCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGATGCTCACTTCTGTGCCGTCTGCAGT 324

Qy       301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
             ||||| ||||| || ||||| || || |||||||| ||||||||||||||||||||||||
Db       325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 384

Qy       361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
             |||||||||||||||||||||||||| |||| |||||||| || ||||||||||| ||
Db       385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA 444

Qy       421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAACTGTTACGAAGTCGGA 480
             || || |||||||| || | ||||||||||||||||||||| ||||||||||| ||| |||
Db       445 GACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 504

Qy       481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
             ||||| |||||||| ||| |||||| || |||||||||||  | || || |||||||||
Db       505 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA 564

Qy       541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCAGCGCCC 600
             ||||||  ||||||| |||| ||      ||| ||||||||||  | || ||| | |||
Db       565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAACAGAACCAGTGGGCACACACCC 624

Qy       601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
             || |||  |||||| |||||| || | |||| |||||||||| ||||||||||| ||| ||
Db       625 CGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCCGAGCAGCTGGTGCTCACCCTGCTG 684

Qy       661 CAGCCTCAGCCCCCCCCATGTGCTCATCAGCCCCCCAGTGGCCCTTCAGCACCCCTCC 720
             || |||||||| ||| ||||||| | || || |||||  |||||||||||||||||||
Db       685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy       721 ATGATGATGTCCCTGACCAAGTTGGGCGACAAGGAGTTGGTACACATGATCAGCTGGCCC 780
             |||||||||||||| || ||| |||| |||||||| |||| ||||||||  ||||||||
Db       745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy       781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCTGCTCTTGGAGAGC 840
             ||||| || || |||||||||||||||||||||||||| |||||||| || ||||||| |||
Db       805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCTGCTCTTGGAAAGC 864

Qy       841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
             || ||||||||||| | ||| |||||||||||||||||| || ||||||||||||||||
Db       865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy       901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
             |||||||||||||||| || ||||||||||||||||||||||| ||||| |||| ||| |||
Db       925 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTGGAAGGGATT 984

Qy       961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
             |||||||||||||||||||||||||||| || ||  || |||| || ||||||||||| ||
Db       985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTGCAG 1044

Qy      1021 CACAAAGAATATCTGTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
             |||||||||||||| ||||| |||||||||| || ||||| ||||||||||||   ||
Db      1045 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG 1104

Qy      1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCGGTG 1140
             |  || || | |||||| || || || ||||||||||| | ||| |  |||||| |||
Db      1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1164

Qy      1141 ACGGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
```

```
             || |||||  |||| |||||||||| | |||||  | ||||| |||||||||||| || |
Db      1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGAATCTCTTCCCAGCAGCAGTCAGTC 1224

Qy      1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
             || |||| |||||||||||||| || || || ||||||||    ||||||||||||||||
Db      1225 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGGCACATCAGTAACAAGGGCATG 1284

Qy      1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
             |||||||||||| |||||||||||||||||||||||||| |||| ||||||||||||||
Db      1285 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTACGACCTGCTGCTGGAG 1344

Qy      1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
             |||||||||| ||| |||||| |||| |||||||| ||| ||||||| | |||| ||
Db      1345 ATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGGTGCTGC 1404

Qy      1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1444
             || ||||||||||||| ||||||||||||||||||||||  ||||||||||||
Db      1405 TCGACAGAGGACAGTAAGAGCAAACAGGGCTCCCAGAACCTCCAGTCTCAGTGA 1458


RESULT 37
AX022460
LOCUS       AX022460               1458 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 6 from Patent EP0945000.
ACCESSION   AX022460
VERSION     AX022460.1  GI:10046070
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT      365 a    373 c    403 g    317 t
ORIGIN

  Query Match             75.3%;  Score 1080.4;  DB 6;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 1.3e-233;
  Matches 1213;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGAGACC 60
              ||||| |||||  |||||||||     |  |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACTTGGAAGGTGGGCCTGTTCGGCAGACT 84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | ||||||||  |||| |||| ||||||||||     ||| ||
Db         85 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| | ||| ||||| |||||||||||||||| ||||||||||||||||||| |||
Db        145 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              |||||||||||| |||||||||||||||||| ||||| ||||| || | |||| | ||||
Db        205 GAACACACCTTGCCTGTAAACAGAGAGACCCTGAAGAGGAAGCTTGGCGGGAGCGGTTGT 264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGC 300
              |||||||||||||| ||||| |  | ||||||||||||||||||||||| || ||||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGATGCTCACTTCTGTGCCGTCTGCAGT 324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || ||||| || || |||||||| ||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||| |||| |||||||| || ||||||||||| ||
Db        385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA 444

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAACTGTTACGAAGTCGGA 480
              || || |||||||| || | ||||||||||||||||||||| ||||||||||| ||| |||
Db        445 GACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 504
```

```
Qy        481 ATGTGAAGTGTGGCTCCCGGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| ||||||| || |||||||||||  | || || ||||||||
Db        505 ATGGTCAAGTGTGGATCCAGGAGAAAGGTGTGGCTACCGAATAGTACGAAGACAGAGA 564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              |||||| |||||| |||| ||       ||| ||||||||||  | || ||| | |||
Db        565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCAGTGGGCACACACCC 624

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || ||  |||||| |||||| || | |||||  ||||||||||| |||||||||| |||
Db        625 CGGGTGAAGGAGCTACTGCTGAACTCTCCTGAGTCCGGACCAGCTGGTGCTCACCCTGCTG 684

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || |||||||| ||| |||||||  | || || |||||   ||||||||||||||||||||
Db        685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy        721 ATGATGATGTCCCTGACCAAGATTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||| || ||| |||| |||||||  |||| ||||||||  ||||||||
Db        745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTCGAGCTCAGCCTGTTCGACCAAGTGCGGGTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||||| |||||||| || |||||||| |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC 864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| | ||| |||||||||||||||||| || ||||||||||||||||
Db        865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              |||||||||||||||| || |||||||||||||||||||||| |||| |||| |||
Db        925 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGATGAGGGCAAGTGCGTGGAAGGCATT 984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||||| || ||  || |||| || |||||||||| ||
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTCCAG 1044

Qy       1021 CACAAAGAATATCTCTGTCTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||||||||| |||| |||||||||| || |||| |||||||||||   ||
Db       1045 CACAAAGAATATCTGTGTCTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG 1104

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |  || || | |||||| || || || |||||||||| | ||| |  ||||| ||||
Db       1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1164

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || ||||  |||| ||:|||||| | |||||  | |||||  |||||||||| ||  |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGATTAGAATCTCTTCCCAGCAGCAGTCAGTC 1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || |||||  |||||||||||| || || || ||||||||   ||||||||||||||||
Db       1225 CGTCTGGCCAACCTCCTGATGCTCCTTTCTCATGTCAGGCACATCAGTAACAAGGGCATG 1284

Qy       1261 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCCGTGTATGACCTGCTGCTGGAG 1320
              |||||||||||||| ||||||||||||||||||||||||| |||||  ||||||||||||||
Db       1285 GAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCCGTGTATGACCTGCTGCTGGAG 1344

Qy       1321 ATGCTGAATGCCCACGTGCTTCGGGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              |||||||||| ||  |||||| |||| |||||||||| ||| ||||||| | |||| ||
Db       1345 ATGCTGAATGCTCACAGCCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGGGTGCTGC 1404

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              || |||||||||||| ||||||||||||||| |||  ||||||||||
Db       1405 TCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAGAACCTCCAGTCTCAGTGA 1458


RESULT 48
MMAJ220
LOCUS       MMAJ220                 1458 bp    DNA     linear   ROD 30-NOV-1997
DEFINITION  Mus musculus mRNA for estrogen receptor subtype beta.
ACCESSION   AJ000220
VERSION     AJ000220.1  GI:2370296
KEYWORDS    estrogen receptor; subtype beta.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Pettersson,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1997) Pettersson K., Department of Medical
            Nutrition, Karolinska Institute, F 60 Novum, S 141 86 Huddinge,
            SWEDEN
REFERENCE   2  (bases 1 to 1458)
  AUTHORS   Pettersson,K., Grandien,K., Kuiper,G.G. and Gustafsson,J.A
  TITLE     The mouse estrogen receptor beta form estrogen response element
            binding heterodimers with the estrogen receptor alpha
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             1..1458
                     /codon_start=1
                     /product="estrogen receptor, subtype beta"
                     /protein_id="CAA03949.1"
                     /db_xref="GI:2370297"
                     /db_xref="SPTREMBL:O35635"
                     /translation="MAFYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPTSGHLSPL
                     ATHCQSSLLYAEPQKSPWCEARSLEHTLPVNRETLKRKLGGSGCASPVTSPSTKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACR
                     LRKCYEVGMVKCGSRRERCGYRIVRRQRSASEQVHCLNKAKRTSGHTPRVKELLNSL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLHDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTARFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDAL
                     VWVISKSFISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHVLRGYKSSISGSQCSSEDSKSKEQSQNLQSQ"
BASE COUNT       365 a    373 c    403 g    317 t
ORIGIN

  Query Match             75.3%;  Score 1080.4;  DB 10;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 1.3e-233;
  Matches 1213;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||||||||    |  |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGGCAGACT 84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCAICGC 120
              |||||||||||||| | ||||||||  |||| |||| ||||||||||    ||| ||
Db         85 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| | |||  |||| ||||||||||||||||| ||||||||||||||| |||
Db        145 CAATCATCGCTTTCTCTATGCAGAACCTCAAAAGAGTCCTGGTGTGAAGCAAGATCACTA 204

Qy        181 GAACACACCTTACCTGTAAACAGACAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||| |||||||||||||| ||||| |||||| || | |||| | ||||
Db        205 GAACACACCTGGCCTGTAAACAGAGAGACCTGAAGAGGAAGCTTGGCGGGAGCGGTTGT 264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGTGCTGTCTGCAGC 300
              |||||||||||||| ||||| |  | ||||||||||||||||||| || |||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGATGCTCACTTCTGTGCCGTCTGCAGT 324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              |||| |||| || |||| || || |||||||| |||||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 384

Qy        361 AAAAGAAGCATTCAAGGACATAATGACTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||| |||| |||||||| || ||||||||| ||
Db        385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA 444

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| ||   |||||||||||||||||||| ||||||||||||| |||
Db        445 GACAACAACCGGCGTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 504
```

```
Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| ||||||| || ||||||||||  | || || ||||||||
Db        505 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGCTACCGAATAGTACCAAGACAGAGA 564

Qy        541 AGTGCCGACGAGCAGCTGCACTCTGCCGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||  ||||||| |||| ||      ||| ||||||||||  | || ||| | |||
Db        565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCACTGGGCACACACCC 624

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACGCTGCTG 660
              || |||  |||||| |||||| || | |||| ||||||||||| ||||||||||| |||
Db        625 CGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCCGAGCAGCTGGTGCTCACGCTGCTG 684

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCACCCGCCCCAGTGCCCCTTCACCCAGCGCTCC 720
              || ||||||| || ||| |||| |||||||  |||| |||||||  ||||||||
Db        685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy        721 ATGATGATCTCCCTGACCAAGTTGGCCGACAACGAGTTGGTACACATGATCAACTGGGCC 780
              |||||||||||||| || ||| |||| |||||||  |||| |||||||  ||||||||
Db        745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||| |||||||| || |||||||| |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC 864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||| | ||| ||||||||||||||| || ||||||||||||||||||||
Db        865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              |||||||||||||||| || |||||||||||||||||||||| ||||| |||| |||
Db        925 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGATGAGGGCAAGTGCGTGGAAGGGATT 984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||| || ||  || ||| || ||||||||||| ||
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTGCAG 1044

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||||||||| ||||| |||||||||| || ||||| ||||||||||||   ||
Db       1045 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG 1104

Qy       1081 GTCACAGCGACCCAAGATGGTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACCCCGTG 1140
              |  || || | |||||| || || || |||||||||||| | ||| |  ||||||| |||
Db       1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACCCACTG 1164

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||| | |||||  | |||| |||||||||| ||  |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGAATCTCTTCCCAGCAGCAGTCAGTC 1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              || |||||  |||||||||||||| || || || |||||||     |||||||||||||
Db       1225 CGTCTGGCCAACCTCCTGATGCTCCTTTCTCATGTCAGGCACATCAGTAACAAGGGCATG 1284

Qy       1261 GAACATCTGCTCAGCATGAAGTGCAAAAAGTGGGTCCCACTGTATGACCTGCTGCTGGAG 1320
              |||||||||||||| ||||||||||||||||||||||||| |||||  ||||||||||||||
Db       1285 GAACATCTGCTCAGCATGAAGTGCAAAAAGTGGGTCCCGGTGTACGACCTGCTGCTGGAG 1344

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              |||||||||| |||  |||||| |||| |||||||| ||| |||||| | |||| ||
Db       1345 ATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGGGTGCTGC 1404

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
                || ||||||||||||| ||||||||||||||||||||||||  ||||||||||||
Db       1405 TCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAGAACCTCCAGTCTCAGTGA 1458


RESULT 39
RNAJ2603
LOCUS       RNAJ2603                1512 bp    DNA     linear   ROD 07-APR-1998
DEFINITION  Rattus norvegicus mRNA for oestrogen receptor beta2 (clone
            pCR.ERbeta2).
ACCESSION   AJ002603
VERSION     AJ002603.1  GI:2632168
KEYWORDS    oestrogen receptor beta.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1512)
  AUTHORS   Aldridge,T.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) Aldridge T.C., Molecular Endocrinology
            Group, ZENECA Central Toxicology Laboratory, Alderley Park,
            Macclesfield, Cheshire SK10 4TJ, ENGLAND
REFERENCE   2  (bases 1 to 1512)
  AUTHORS   Aldridge,T.C.
  TITLE     Tissue specific responses to estrogen: an explanation based on
            differential activation of multiple estrogen receptors with
            different estrogen response elements
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1..1512
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone="pCR.ERbeta2"
                     /tissue_type="Prostate"
     gene            1..1512
                     /gene="ERbeta"
     CDS             1..1512
                     /gene="ERbeta"
                     /note="Splice variant of oestrogen receptor beta. Generated
                     by RT-PCR from total RNA. The sequence contains two amino
                     acid changes from the previously published ERbeta
                     sequence, along with a 54 base pair (18aa) insertion in
                     the ligand binding domain."
                     /codon_start=1
                     /evidence=experimental
                     /product="oestrogen receptor beta2"
                     /protein_id="CAA05632.1"
                     /db_xref="GI:2632169"
                     /db_xref="SPTREMBL:O35785"
                     /translation="MTFYSPAVMNYSVPGSTSNLDGGPVRQSTSPNVLWPTSGHLSPI
                     AIHQQSSLIYAEPQKSPWCEAKSLEHTLPVNPELLKPKLSGSSCASPVTSPNAKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LRKCYEVGMVKGSRFFKQYRFVPPQKSSSFQVHCTSKAKRNSGHAPRVKELLLSTL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     ISLFDQVRLIESCWMEVLMVCLMWRSIDHPGKLIFAPDLVIDRSSFDPHWHVAQMKSA
                     APRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQE
                     AFSSRKITHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLL
                     SMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSECSSIEDSKNKESSQNLQSQ"
BASE COUNT      384 a    390 c    415 g    323 t
ORIGIN

  Query Match          72.2%;  Score 1035.6;  DB 19;  Length 1512;
  Best Local Similarity   82.3%;  Pred. No. 1.7e-223;
  Matches 1225;  Conservative    0;  Mismatches  209;  Indels   54;  Gaps    1;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||| ||||    || |||| |||| ||||||||||  || |||| |
Db         25 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACAGAGC 84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | |||||||  |||||||||| |||||||| |  |||| ||
Db         85 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| |  || |||| ||||||||||||||| ||||||||||||||| |||
Db        145 CAATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy        181 GAACACACCTTACCTGTAAACAGAGACACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||||| |||||| |||||||| | ||||
Db        205 GAGCACACCTTACCTGTAAACAGAGACACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGT 264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||| |||||     |||||||||||||||||||||||| |||||||||
```

```
Db        265 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCGCCGTCTGCAGC 324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || |||| || || || ||||| ||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTT 384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||||||||| |||||||| || |||||||||| ||
Db        385 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 444

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| || ||||||||||||||||||||| |||||||| ||||| |||
Db        445 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 504

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| ||||||| || ||||||| | |||||||||| | |||||||| |||||
Db        505 ATGGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGA 564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCCGGCAAGGCCAAGAGAAGTGGCGGGCCACGCGCCC 600
              ||| |   ||||||| | |||||     |||| ||||||||||  || || || || |||
Db        565 AGTCCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCC 624

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| ||||||  | || |||| || ||||| || ||||||||||||||||
Db        625 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCATAGCAACTGGTGCTCACCCTCCTG 684

Qy        661 GAGGCTGATCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCC 720
              || |||| || ||| |||||||| | |||| ||||   ||||||||||||||||||||||
Db        685 GAAGCTGAACCACCCAATGTGCTGGTCAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy        721 ATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGCTGGTCACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||  |||| |||||||  |||||||||
Db        745 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||| |||||||| |||||||| || |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGC 864

Qy        841 TGTTCGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||| ||||| |||| ||||||||||||| || ||||||||||||||||||
Db        865 TGCTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGAC ---------- -- ------ ------  930
              ||||| || ||||| || || ||||||||
Db        925 CTCATTTTCGCTCCCGACCTCGTTCTGGACAGGTGCTCAGAAGACCCTCACTGGCACGTT 984

Qy        931 -   -   -     AGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAA 966
                            |||||||||||| |||||||||| |||||||||||
Db        985 GCGCAGATGAAGAGTGCTGCCCCAAGGGATGAGGGGAAGTGCGTAGAAGGGATTCTGGAA 1044

Qy        967 ATCTTTGACATGCTCCTGGCAACTACTTCAAGTTTCGAGAGTTAAAACTCCAACACAAA 1026
              |||||||||||||||||||| || || |||||||| || ||||||||||||| |||||
Db       1045 ATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCGTGAGTTAAAACTCCAGCACAAG 1104

Qy       1027 GAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCCTGGTCACA 1086
              || |||||||||| |||||||||||||| |||| ||||||||||||||  ||| | |
Db       1105 GAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTGGCTTCT 1164

Qy       1087 GCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGAT 1146
              || | |||||| || || || |||||||||||| | ||| | |||||||| ||||| |||
Db       1165 GCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACTTACTGAACGCGGTGACAGAT 1224

Qy       1147 GCTTTGGTTTGGGTCATTCCCAAGACCCCCATCTCCTCCCAGCAGCAATCCATGCGCCTG 1206
              ||  |||| |||||||||| |||| || ||||||||||||||||| ||  | || |||
Db       1225 GCCCTGGTCTGGGTGATGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCGACTG 1284

Qy       1207 GCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACAT 1266
              || |||||||||||||| || || ||||||||||||    ||||||||||||||||||||
Db       1285 GCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATGGAACAT 1344

Qy       1267 CTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTG 1326
              ||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1345 CTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGATGCTG 1404

Qy       1327 AATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCA 1386
              ||||| ||| |||||| |||| ||||||||| ||| ||||||| |||||||| |  ||
Db       1405 AATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGCTCAACA 1464

Qy       1387 GAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              |||||||||||| | ||||||| |||||||||||| ||||||||||||
Db       1465 GAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGA 1512


RESULT 40
AF042059
LOCUS       AF042059                1675 bp    mRNA    linear   ROD 22-JAN-1998
DEFINITION  Rattus norvegicus estrogen receptor beta2 (ERbeta) mRNA, complete
            cds.
ACCESSION   AF042059
VERSION     AF042059.1  GI:2801692
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1675)
  AUTHORS   Petersen,D.N., Tkalcevic,G.T., Koza-Taylor,P.H., Turi,T.G. and
            Brown,T.A.
  TITLE     Identification of estrogen receptor beta2, a functional variant of
            estrogen receptor beta expressed in normal rat tissues
  JOURNAL   Endocrinology (1998) In press
REFERENCE   2  (bases 1 to 1675)
  AUTHORS   Brown,T.A. and Petersen,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1998) Molecular Sciences, Pfizer Central
            Research, Eastern Point Road, Groton, CT 06340, USA
FEATURES             Location/Qualifiers
     source          1..1675
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1..1675
                     /gene="ERbeta"
     CDS             1..1512
                     /gene="ERbeta"
                     /note "steroid hormone receptor; spliced variant of ERbeta
                     containing an insertion within the ligand binding domain"
                     /codon_start=1
                     /product "estrogen receptor beta2"
                     /protein_id="AAB97425.1"
                     /db_xref="GI:2801693"
                     /translation="MTPYSPAVMNYGVPGSTGNLLEGRGVRQGISPNVLWPTSGRLSPL
                     AIHCQSSLGYAEPQKSPWCEAKSLEHTLPVNKETLKKKLSGSSCASPVTSPNAKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LKKCYEVGMVKGGSRKERGGYRTVRRQRSSSEQVPGLSKAKRNGGHAPKVKELLTSIL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMISWAKKIPGFVE
                     LSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRSSFEGQSHGHVAGMSSA
                     APRDEGKCVEGLLEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQE
                     AESSRKLTHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLL
                     SMKCKNVVPVYDLLLEMLNAHILEGYDSSTSQSPQSSSLEDSKNKPSSQNLQSQ"
     variation       80
                     /gene="ERbeta"
                     /note="encodes leucine residue in GenBank Accession Number
                     U57439"
                     /replace="t"
     variation       313
                     /gene="ERbeta"
                     /note="encodes proline residue in GenBank Accession Number
                     U57439"
                     /replace="c"
BASE COUNT      417 a    434 c    464 g    360 t
ORIGIN

  Query Match          72.2%;  Score 1035.6;  DB 10;  Length 1675;
  Best Local Similarity   82.3%;  Pred. No. 1.8e-223;
```


Matches 1225; Conservative 0; Mismatches 209; Indels 54; Gaps 1;
Search completed: June 16, 2003, 07:38:46
Job time : 3807.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 03:20:57 ; Search time 139.67 Seconds
(without alignments)
505.642 Million cell updates/sec

Title: US-08-826-361A-25
Perfect score: 2805
Sequence: 1 MDIKNSPSSLNSPSSYNCSQ.....ECSPAEDSKSKFGSQNPQSQ 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1983.DAT:*
2: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA_new.DAT:*
22: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA_upd.DAT:*
23: /SIDS2/qcqdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2805	100.0	530	18	AAW33215	Human oestrogen re
2	2805	100.0	530	20	AAW97858	Human oestrogen re
3	2805	100.0	530	22	AAE10457	Human oestrogen rec
4	2805	100.0	530	22	AAU27322	Human Oestrogen re
5	2805	100.0	530	22	AAB60649	Human oestrogen re
6	2805	100.0	530	23	AAB47825	Full length estrog
7	2805	100.0	548	20	AAY07270	Human oestrogen re
8	2805	100.0	793	23	ABB80756	Amino acid sequenc
9	2554	91.1	485	18	AAW14724	Human Oestrogen re
10	2535	90.4	485	20	AAY01597	An estrogen recept
11	2522	89.9	477	18	AAW33212	Human oestrogen re
12	2522	89.9	477	23	AAB47834	Estrogen receptor.
13	2502	89.2	549	20	AAW98125	Mouse oestrogen re
14	2500	89.1	549	20	AAY04434	Murine mER beta-1
15	2481	88.4	567	20	AAW98128	Mouse complete oes
16	2457	87.6	560	20	AAY04433	Murine mER beta-3
17	2291	81.7	485	18	AAW14723	Rat oestrogen rece
18	2291	81.7	485	20	AAY01596	An estrogen recept
19	2267	80.8	485	18	AAW14725	Mouse oestrogen re
20	2245.5	80.1	510	20	AAY04435	Murine mER beta-2
21	2245.5	80.1	510	20	AAW98126	Mouse oestrogen re
22	2231.5	79.6	474	21	AAY51946	Bovine ER-beta pro
23	2198	78.4	416	18	AAW33213	Human oestrogen re
24	2198	78.4	416	23	AAB47837	ER splice variant
25	2198	78.4	418	18	AAW33214	Human oestrogen re
26	2198	78.4	418	23	AAB47836	ER splice variant
27	1746	62.2	414	20	AAW98127	Rat oestrogen rece
28	1746	62.2	730	20	AAY04436	Rat rER beta-4 clo
29	1235.5	44.0	581	23	ABB83791	Chemidophorus unip
30	1235.5	44.0	595	22	AAG84509	Human oestrogen re
31	1233.5	44.0	595	20	AAY21626	Ligand binding dom
32	1233.5	44.0	595	22	AAG84505	Human oestrogen re
33	1233.5	44.0	595	22	AAG84506	Human oestrogen re
34	1233.5	44.0	595	22	AAG84513	Human oestrogen re
35	1233.5	44.0	595	23	ABB76378	Human nuclear oest
36	1231.5	43.9	595	22	AAG84512	Human oestrogen re
37	1230.5	43.9	595	22	AAG84507	Human oestrogen re
38	1226.5	43.7	595	22	AAG84508	Human oestrogen re
39	1226.5	43.7	595	22	AAG84514	Human oestrogen re
40	1224.5	43.7	595	20	AAY33506	Human estrogen rec
41	1224.5	43.7	595	22	AAG84510	Human oestrogen re
42	1224.5	43.7	595	22	AAG84511	Human oestrogen re
43	1224.5	43.7	595	23	ABB81783	Human oestrogen re
44	1224.5	43.7	595	23	ABB82265	Human oestrogen re
45	1224.5	43.7	595	23	AAG68251	Human oestrogen re
46	1216.5	43.4	587	23	ABB81600	Caiman crocodilus
47	1216.5	43.4	595	8	AAP70543	Human oestrogen re
48	1214.5	43.3	582	22	AAC78227	Lepomis centrarchi
49	1199	42.7	595	23	AAE19430	Medaka fish oestro
50	1187	42.3	457	22	AAB61499	Protein encoded by
51	1186.5	42.3	422	22	AAB61498	Protein encoded by
52	1186	42.3	692	23	AAM48333	Oestrogen receptor.
53	1185	42.2	595	21	AAB20897	Oryzias lapites oe
54	1185	42.2	600	21	AAB20898	Oryzias lapites oe
55	1181	42.1	233	23	AAB47843	Ligand binding dom
56	1176	41.9	233	18	AAW33211	Human oestrogen re
57	1168.5	41.7	506	22	AAG78226	Lepomis centrarchi
58	1126.5	40.2	229	22	AAB98830	Human ERbeta-genis
59	1104	39.4	228	22	AAB98831	Rat ERbeta Raloxif
60	1013	36.1	344	9	AAP89925	Sequence of the hu
61	813.5	29.0	910	23	AAB47902	E2C-ER-1-ER-VP64.
62	813.5	29.0	922	23	AAB47903	E2C-ER-11-ER-VP64.
63	811	28.9	349	23	AAU98699	Endocrine disrupti
64	810	28.9	774	20	AAY21544	SSR-LBD fusion pol
65	784.5	28.0	740	23	AAM51591	Fused ER protein.
66	771	27.5	480	22	AAB36684	Mammalian two-hybr
67	770	27.5	244	21	AAB26780	Human oestrogen re
68	770	27.5	244	21	AAB26784	Oestrogen receptor
69	768	27.4	241	23	AAE17864	Estrogen receptor
70	763	27.2	651	19	AAW47583	Multi-chimeric tra
71	763	27.2	660	23	ABB76379	Cre recombinase-hu
72	762	27.2	651	19	AAW48357	Multi-chimeric tra
73	760.5	27.1	241	21	AAB26782	Oestrogen receptor
74	760	27.1	660	23	ABB76381	Cre recombinase-hu
75	751	26.8	660	23	ABB76380	Cre recombinase-hu
76	746.5	26.6	264	22	AAB80952	Equine oestrogen r
77	745.5	26.6	245	21	AAB26783	Oestrogen receptor
78	745.5	26.6	245	23	AAU75931	Oestrogen receptor
79	713	25.4	229	21	AAB26789	Oestrogen receptor
80	648	23.1	435	19	AAW61192	Steroid hormone re
81	648	23.1	435	21	AAY82263	Human oestrogen re
82	648	23.1	458	20	AAY03837	Human nuclear rece
83	648	23.1	458	21	AAB09965	Human ERRgamma pro

84	648	23.1	458	21	AAY83823	Human nuclear rece
85	643	22.9	508	20	AAY29467	Human SBPERR4 prot
86	633	22.6	433	21	AAB12970	Human oestrogen re
87	632.5	22.5	433	9	AAP80931	Sequence of human
88	631.5	22.5	433	21	AAB09966	Human ERRbeta prot
89	630	22.5	500	20	AAY03836	Human nuclear rece
90	630	22.5	500	21	AAY83822	Human nuclear rece
91	619	22.1	374	22	AAU02986	Angiotensin conver
92	619	22.1	382	22	AAU02958	Angiotensin conver
93	600	21.4	552	22	AAM39411	Human polypeptide
94	597.5	21.3	521	9	AAP80930	Sequence of human
95	597	21.3	486	23	ABB05347	Human nuclear horm
96	585.5	20.9	556	22	AAM41197	Human polypeptide
97	576	20.5	418	21	AAY83824	Human nuclear rece
98	571	20.4	418	20	AAY03838	Human nuclear rece
99	554	19.8	101	22	ABB43048	Peptide #10544 enc
100	554	19.8	101	22	ABB26221	Protein #8220 enco
101	554	19.8	101	22	AAM63943	Human brain expres
102	554	19.8	101	22	AAM76761	Human bone marrow
103	554	19.8	101	22	AAM36868	Peptide #10905 enc
104	554	19.8	101	23	ABG45934	Human peptide enco
105	527.5	18.8	484	22	ABB63251	Drosophila melanog
106	504	18.0	99	22	AAM37173	Peptide #11210 enc
107	460.5	16.4	525	20	AAY21625	Ligand binding dom
108	459	16.4	446	12	AAR11837	H2RIIBP nuclear ho
109	459	16.4	446	16	AAR72484	H-2RIIBP. Not spe
110	458.5	16.3	533	14	AAR39468	hRXR-beta1. Homo
111	458.5	16.3	533	16	AAR72483	Human H-2RIIBP. H
112	458.5	16.3	533	23	AAU78297	Human Retinoid X R
113	454	16.2	448	14	AAR39467	mRXR-beta. Mus mu
114	454.5	16.2	510	14	AAR39469	hRXR-beta2. Homo
115	452	16.1	463	12	AAR13800	Mouse RXR gamma re
116	452	16.1	463	14	AAR39471	mRXR-gamma. Mus m
117	451	16.1	933	20	AAY21621	Ligand binding dom
118	451	16.1	933	21	AAY97297	Human progesterone
119	449.5	16.0	984	9	AAP80927	Sequence of the hu
120	448	16.0	849	23	AAG68248	Fused androgen rec

ALIGNMENTS

RESULT 1
AAW83215
ID AAW83215 standard; Protein; 530 AA.
XX
AC AAW83215;
XX
DT 20-APR-1998 (first entry)
XX
DE Human oestrogen receptor protein #2.
XX
KW oestrogen receptor protein; steroid; alternative splicing; estradiol;
KW estone; estriol; screening.
XX
OS Homo sapiens.
XX
PN EP798378-A2.
XX
PD 01-OCT-1997.
XX
PF 25-MAR-1997; 97EP-0200903.
XX
PR 22-NOV-1996; 96EP-0203284.
PR 26-MAR-1996; 96EP-0200820.
XX
PA (ALKO) AKZO NOBEL NV.
XX
PI Dijkema R, Mosselman S;
XX
DR WPI; 1997-473188/44.
DR N-PSDB; AAT88415.
XX
FT DNA encoding estrogen receptor useful in screening assay to
FT identify novel ligands or hormonal analoques
XX
PS Claim 4; Page 35-37; 45pp; English.
XX
CC This sequence represents a novel oestrogen binding protein isolated from
CC human testis cDNA In order to study upstream translation-initiation
CC codons using 5' RACE PCR technology. This receptor is able to bind and
CC be activated by estradiol, estone and estriol, can be used in a screening
CC assay for the identification of new drugs e.g. novel ligands or hormonal
CC analoques.
XX
SQ Sequence 530 AA;

Query Match 100.0%; Score 2805; DB 18; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.4e-258;
Matches 530, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRKQKSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NDYICPAINQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRKQKSADEQLH 240

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVIDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DSSRKLAHLLNAVIDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
```

RESULT 2
AAW97858
ID AAW97858 standard; Protein; 530 AA.
XX
AC AAW97858;
XX
DT 07-JUN-1999 (first entry)
XX
DE Human oestrogen receptor-beta.
XX
KW oestrogen receptor-beta; hER-beta; human.
XX
OS Homo sapiens.
XX
PN WO9907847-A1.
XX
PD 18-FEB-1999.
XX

PF 20-JUL-1998; 98WO-US14944.
XX
PR 05-AUG-1997; 97US-0906365.
XX
PA (AMHP) AMERICAN HOME PROD CORP.
XX
PI Bhat RA, Henderson RA, Hsiao C, Karathanasis SK;
XX
DR WPI; 1999-167424/14.
DR N-PSDB; AAX24364.
XX
PT Human oestrogen receptor-beta coding sequence - useful in the
PT production of human oestrogen receptor-beta and identification of
PT human oestrogen receptor-beta interactive compounds
XX
PS Claim 17; Page 42-44; 49pp; English.
XX
CC This polypeptide comprises a full-length human oestrogen receptor-beta
CC (hER-beta), including 45 previously unknown N-terminal amino acid
CC residues that are believed to contribute to the transcription
CC activation function of the receptor. The amino acid sequence was
CC deduced from the nucleotide sequence (see AAX24364) of an isolated
CC cDNA clone. hER-beta is selectively expressed in the thymus,
CC spleen, ovary and testes. The invention encompasses hER-beta
CC polynucleotides and polypeptides, particularly peptides which
CC include residues 1-45 of hER-beta. The invention also provides
CC expression systems in which transcriptionally active hER-beta or
CC fragments can be produced, as well as screening methods for
CC identifying hER-beta agonists and antagonists (including
CC tissue-specific oestrogens and anti-oestrogens), and hER-beta
CC co-activators and inhibitors.
XX
SQ Sequence 530 AA;

Query Match 100.0%; Score 2805; DB 20; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.4e-258;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
Db 1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy 61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
Db 61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120

Qy 121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
Db 121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy 181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVPRQRSADEQLH 240
Db 181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSPPERCGYRLVPRQRSADEQLH 240

Qy 241 CACKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
Db 241 CACKAKRSGGHAPRVRELLLDALSPEQIVITLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy 301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
Db 301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy 361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
Db 361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy 421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
Db 421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy 481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
Db 481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530

RESULT 3
AAE10457
ID AAE10457 standard; Protein; 530 AA.
XX
AC AAE10457;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human estrogen receptor (ER)-beta protein.
XX
KW Estrogen receptor-beta; dietary compound; environmental pollutant;
KW red wine; ER-beta mediated cell growth inhibition modulator; human.
XX
OS Homo sapiens.
XX
PN WO200169262-A1.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-US08276.
XX
PR 15-MAR-2000; 2000US-189605P.
XX
PA (UYMA-) UNIV MASSACHUSETTS.
XX
PI Ho S;
XX
DR WPI; 2001-596920/67.
DR N-PSDB; AAD18442.
XX
PT In vitro screening for modulator of estrogen receptor-beta-mediated
PT cell growth inhibition by contacting cell having functional ER-beta
PT protein with candidate compound and detecting ER beta regulated gene
PT expression -
XX
PS Example 3; Fig 7A; 49pp; English.
XX
CC The invention relates to in vitro screening for modulator of estrogen
CC receptor-beta (ER-beta)-mediated cell growth inhibition. The method
CC involves contacting mammalian cell having functional ER-beta protein
CC with candidate compound and detecting increase/decrease of ER-beta-
CC regulated gene expression in presence of candidate compound compared
CC to expression in absence of the compound. The method is useful for
CC identifying a compound that modulates ER-beta-mediated cell growth
CC inhibition. It is useful for testing dietary compounds, e.g. red wine,
CC for the presence of estrogen mimics that trigger ER-beta-mediated
CC processes, thus affecting physiological function positively or
CC negatively and for testing environmental pollutants for the presence
CC of estrogen mimics that may pose health risks involving ER-beta-
CC mediated processes. The present sequence is human ER-beta protein.
XX
SQ Sequence 530 AA;

Query Match 100.0%; Score 2805; DB 22; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.4e-258;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
Db 1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy 61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
Db 61 NVTNLEGGPGRQTTSPNVIWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120

Qy 121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
Db 121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy 181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240

```
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
```

RESULT 4
AAU27322
ID AAU27322 standard; Protein; 530 AA.
XX
AC AAU27322;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human Oestrogen receptor beta.
XX
KW Human; Oestrogen receptor beta; ERbeta; SNP; chromosome 6q.25.1;
KW single nucleotide polymorphism; cardiovascular disease;
KW autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;
KW osteoarthritis; osteoporosis; breast cancer; endometrial cancer.
XX
OS Homo sapiens.
XX
PN WO200162793-A2.
XX
PD 30-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05360.
XX
PR 22-FEB-2000; 2000US-0183755.
PR 24-JAN-2001; 2001US-0768185.
XX
PA (PEKE) PE CORP NY.
XX
PI Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES,
XX
DR WPI; 2001-582041/65.
DR N-PSDB; AAS43105, AAS43104.
XX
PT Oestrogen receptor gene and protein polymorphisms useful for diagnosis
PT of individuals at risk of developing bone disorders -
XX
PS Disclosure; Figure 4; 245pp; English.
XX
CC The invention relates to a novel isolated peptide comprising or
CC consisting of an amino acid sequence selected from an amino acid sequence
CC of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
CC 10 amino acids), antibodies against them, nucleic acids encoding
CC them (including vectors for transforming cells). The gene for human
CC ERbeta is located on chromosome 6q.25.1. The variants are encoded
CC by single nucleotide polymorphisms (SNP). The variant peptides and
CC proteins can be used in assays to determine the biological
CC activity of the protein, to raise antibodies, as a reagent in assays
CC designed to quantitatively determine levels of the protein in
CC biological fluids, to identify compounds that modulate receptor
CC activity and to screen compounds for the ability to stimulate or
CC inhibit interaction between the receptor protein and a target molecule
CC that normally interacts with the receptor protein e.g. oestrogen.
CC The antibody can be used to isolate the protein, to assess expression in
CC disease states e.g. cardiovascular disease and autoimmune disease (e.g.
CC systemic lupus erythematosus, arthritis, rheumatism and osteoarthritis),
CC osteoporosis, breast cancer and endometrial cancer. In addition
CC the antibodies can be used in pharmacogenomic analysis and inhibiting
CC protein function, e.g. blocking the binding of the oestrogen receptor
CC protein to a binding partner such as a ligand. The nucleic acids
CC encoding the proteins can be used as probes, primers, chemical
CC intermediates and in biological assays. The present sequence represents
CC human ERbeta.
XX
SQ Sequence 530 AA;

```
  Query Match             100.0%;  Score 2805;  DB 22;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-258;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEAPSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEAPSLEHTLPVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
```

RESULT 5
AAB60649
ID AAB60649 standard; Protein; 530 AA.
XX
AC AAB60649;
XX
DT 04-MAY-2001 (first entry)
XX
DE Human oestrogen receptor beta (ER-beta), SEQ ID NO:1.
XX
KW Human oestrogen receptor beta, ER beta, labelled hormone receptor;
KW optical signal; ligand binding; hormonal disorder;
KW sex hormone associated, ovulation, drug screening, drug discovery.
XX
OS Homo sapiens.
XX
PN WO200107919-A1.

```
XX
PD   01 FEB 2001
XX
PF   24-JUL-2000; 2000WO-IP04930
XX
PR   23-JUL-1999;   99JP-0209860.
PR   31-MAY-2000; 2000JP-0163475.
PR   31-MAY-2000; 2000JP-0163476.
XX
PA   (OLYU ) OLYMPUS OPTICAL CO LTD.
XX
PI   Sakamoto H,  Kato N;
XX
DR   WPI; 2001-168581/17.
DR   N-PSDB; AAF59897.
XX
PT   Determination of the interaction of a substance for investigation and
PT   diagnosis of hormonal disorders using an optically labelled receptor
PT   protein -
XX
PS   Claim 26; Page 68-72; 75pp; Japanese.
XX
CC   The invention relates to a method for examining the interaction of a
CC   test substance with a hormone receptor protein. The method involves
CC   contacting a test substance with a hormone receptor protein (e.g., an
CC   oestrogen receptor) which is labelled with an optical marker (e.g., a
CC   fluorescent protein), where the marker is capable of undergoing a change
CC   in its optical properties when the receptor binds a ligand. The optical
CC   signal generated in the presence of the test compound is then compared
CC   with that generated in the absence of the test compound to determine
CC   whether the test compound is interacting with the hormone receptor
CC   protein. The invention also encompasses the labelled hormone receptor
CC   protein, the gene encoding it, and vectors and host cells comprising
CC   such genes. The method of the invention is used for the investigation
CC   and diagnosis of hormonal disorders, especially those associated with
CC   sex hormones e.g., suppression of ovulation. The present sequence
CC   represents human oestrogen receptor beta (ER-beta).
XX
SQ   Sequence   530 AA;

  Query Match             100.0%;  Score 2805;  DB 22;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-258;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYPLVPRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYPLVPRQRSADEQLH 240

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530


RESULT 6
AAB47835
ID   AAB47835 standard; Protein; 530 AA.
XX
AC   AAB47835;
XX
DT   25-MAR-2002  (first entry)
XX
DE   Full length estrogen receptor.
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER alpha; ER beta; exon 8;
KW   estradiol; nuclear receptor; progesterone receptor.
XX
OS   Homo sapiens.
XX
PN   EP1162264-A2.
XX
PD   12-DEC-2001.
XX
PF   25-MAR-1997; 2001EP-0202021.
XX
PR   26-MAR-1996;   96EP-0200820.
PR   22-NOV-1996;   96EP-0203284.
PR   25-MAR-1997;   97EP-0200903.
XX
PA   (AKZO ) AKZO NOBEL NV.
XX
PI   Mosselman S,  Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   N-PSDB; AAI72144.
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor -
XX
PS   Example A; Page 26-28; 35pp; English.
XX
CC   This sequence shows the full length novel estrogen receptor (ER). The
CC   additional N-terminal sequences of this sequence, compared to the ER
CC   sequence given in AAB47834, were generated using RACE PCR. The gene
CC   encoding this new ER is located on chromosome 14 and has a different
CC   tissue distribution from classical ER. This ER also has two orphan ER's,
CC   ER alpha and ER beta.  These orphan receptors have estrogen receptor
CC   related structure but do not appear to able to bind estradiol or other
CC   ER ligands. The DNA binding domain (DBD) and ligand binding domain (LBD)
CC   from this ER may be used in the chimeric receptor of the invention
CC   which also has an N-terminal domain.  The chimeric receptor, or DNA
CC   encoding it, is useful in a screening assay for identification of new
CC   drugs. Similar chimeric receptors comprising the LBD of the new ER,
CC   and also comprising the DBD and an N-terminal domain derived from
CC   another nuclear receptor e.g., progesterone receptor, are useful for the
CC   screening of compounds to identify new ligands or hormone analogs which
CC   are able to activate the new ER. Chimeric receptors comprising a DBD
CC   of the new ER, and LBD and an N-terminal domain derived from
CC   another nuclear receptor, can be used to identify new ligands or
CC   hormone analogs for the nuclear receptors.
XX
SQ   Sequence   530 AA;

  Query Match             100.0%;  Score 2805;  DB 23;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-258;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
```

RESULT 7
AAY07270
ID AAY07270 standard; Protein; 548 AA.
XX
AC AAY07270;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human oestrogen receptor.
XX
KW Human; oestrogen receptor; ligand; bone resorption; metabolic disorder;
KW cardiovascular disease; cancer; central nervous system; breast; uterine;
KW osteoporosis; ovarian; prostate; diabetes; Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN WO9912961-A1.
XX
PD 18-MAR-1999.
XX
PF 04-SEP-1998; 98WO-US18577.
XX
PR 20-MAR-1998; 98GB-0006042.
PR 08-SEP-1997; 97US-0058271.
PR 30-SEP-1997; 97US-0060520.
PR 30-OCT-1997; 97GB-0022884.
XX
PA (MERI) MERCK & CO INC.
XX
PI Wilkinson H;
XX
DR WPI; 1999-229222/19.
DR N-PSDB; AAX34309.
XX
PT Estrogen receptor
XX
PS Claim 1; Fig 1; 42pp; English.
XX
CC This sequence represents a human oestrogen receptor. The receptor can
CC be used to identify ligands that bind to human oestrogen receptor. The
CC ligands can be used in a method for preventing or treating an oestrogen
CC receptor mediated disease or condition, such as abnormal bone resorption,
CC a cardiovascular disease, cancer, metabolic disorders, or central nervous
CC system disorders. The ligand is especially used to treat osteoporosis,
CC breast, uterine, ovarian or prostate cancer, diabetes or Alzheimer's
CC disease.
XX
SQ Sequence 548 AA;

Query Match 100.0%; Score 2805; DB 20; Length 548;
Best Local Similarity 100.0%; Pred. No. 1.5e-258;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 78

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 138

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 198

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 258

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 318

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 378

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 438

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 498

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 548
```

RESULT 8
ABB80756
ID ABB80756 standard; Protein; 794 AA.
XX
AC ABB80756;
XX
DT 23-SEP-2002 (first entry)
XX
DE Amino acid sequence of a fusion GFP/hER beta protein.
XX
KW Fluorometric analyzer; laser; fluorescence; GFP; ER beta; human; gene;
KW green fluorescent protein; estrogen receptor beta; fusion protein.
XX
OS Synthetic.
OS Aequorea victoria.
OS Homo sapiens.
XX
PN WO200248693-A1.
XX

```
PD   20-JUN-2002.
XX
PF   14-DEC-2001; 2001WO-JP10998.
XX
PR   14-DEC-2000; 2000JP-0388427.
PR   30-JAN-2001; 2001JP-0022105.
XX
PA   (OLYU ) OLYMPUS OPTICAL CO LTD.
XX
PI   Kato N,  Sakamoto H;
XX
DR   WPI; 2002-528176/56
DR   N-PSDB; ABN86125.
XX
PT   Autocorrection function-embedded confocal optics-based fluorometric
PT   analyzers for studying behaviors of fluorescence-labeled molecules
PT   particularly intracellular biological molecules like proteins at
PT   molecular level -
XX
PS   Example 2; Page 98-105; 109pp; Japanese
XX
CC   The invention relates to a fluorometric analyzer that comprises a laser
CC   light source, an optical system for focusing the light beam from laser
CC   light source to a sample to form a confocal region, another optical
CC   system for focusing fluorescence from the sample, a light detector for
CC   measuring intensity, and a recording means to record the variation of
CC   intensity in the measured fluorescence. The analyzers are for studying
CC   behaviors of fluorescence-labeled molecules particularly intracellular
CC   biological molecules like proteins at molecular level, e.g. protein
CC   functions and interactions. Such analyzers are stable and convenient to
CC   operate, thereby enabling easy performance of fluorescence correlation
CC   spectroscopy, fluorescence intensity distribution analysis and
CC   fluorescence intensity multiple distribution analysis. The present
CC   sequence represents the amino acid sequence of a fusion protein
CC   comprising a GFP (green fluorescent protein) and hER (human estrogen
CC   receptor) beta protein.
XX
SQ   Sequence   793 AA;

  Query Match             100.0%;  Score 2805;  DB 23;  Length 793;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-258;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0,  Gaps    0,

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 323

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 383

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 443

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREPCGYPLVPRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREPCGYPLVPRQRSADEQLH 503

Qy        241 CAGKAKRSGGHAPRVRELLIDAISPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 CAGKAKRSGGHAPRVRELLIDAISPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 563

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 623

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        624 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 683

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        684 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 743

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 793


RESULT 9
AAW14724
ID   AAW14724 standard; Protein; 485 AA.
XX
AC   AAW14724;
XX
DT   08-JUN-1997  (first entry)
XX
DE   Human oestrogen receptor beta (ER-beta).
XX
KW   Orphan receptor; oestrogen receptor beta; ER-beta;
KW   nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW   ovary cancer; cardiovascular disease; osteoporosis;
KW   environment; pollutant.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          104..169
FT                   /label= DNA-binding_domain
FT   Domain          260..457
FT                   /label= Ligand-binding_domain
XX
PN   WO9709348-A2.
XX
PD   13-MAR-1997.
XX
PF   09-SEP-1996;   96WO-EP03933.
XX
PR   08-MAY-1996;   96GB-0009576.
PR   08-SEP-1995;   95GB-0018272.
PR   15-MAR-1996;   96GB-0005550.
PR   11-APR-1996;   96GB-0007532.
XX
PA   (KARO ) KARO BIO AB.
XX
PI   Enmark E,  Gustafsson J,  Kuiper GG;
XX
DR   WPI; 1997-192842/17.
DR   N-PSDB; AAT62843.
XX
PT   New isolated oestrogen receptor beta - used to develop prods. for
PT   treating e.g. cancers, CNS diseases, osteoporosis or cardiovascular
PT   disease
XX
PS   Claim 1; Fig 13A; 45pp; English.
XX
CC   A novel human orphan receptor (AAW14724) is related to the known
CC   oestrogen receptor ER-alpha, and has been designated ER-beta.
CC   It is an oestrogen receptor-related nuclear receptor.  The ER-beta
CC   amino acid sequence was deduced from a cDNA clone (AAT62843) isolated
CC   from a human testis cDNA library.  Rat, human and mouse ER-beta
CC   (AAW14723-25) can be used to isolate molecules for use in the
CC   treatment of cardiovascular diseases, central nervous system
CC   diseases, osteoporosis, prostate or ovarian cancer or benign
CC   prostatic hyperplasia and to test environmental chemicals for
CC   oestrogenic activity.
XX
SQ   Sequence   485 AA;

  Query Match             91.1%;  Score 2554;  DB 18;  Length 485;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-244;
  Matches  484;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         46 MIPYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 120

Qy        166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 180

Qy        226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 240

Qy        286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 300

Qy        346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 360

Qy        406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 420

Qy        466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 480

Qy        526 NPQSQ 530
              | |||
Db        481 NLQSQ 485
```

RESULT 10
AAY01597
ID AAY01597 standard; Protein; 485 AA.
XX
AC AAY01597;
XX
DT 17-JUN-1999 (first entry)
XX
DE An estrogen receptor beta protein.
XX
KW differential estrogen receptor; ER alpha; ER-beta; activation;
KW API site; therapy.
XX
OS Homo sapiens.
XX
PN WO9911760-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US018030.
XX
PR 04-SEP-1997; 97US-0923708.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Gustafsson J, Kuiper GGJM, Kushner PJ, Nilsson S;
PI Paech K, Scanlan TS, Webb P;
XX
DR WPI; 1999-205174/17.
DR N-PSDB; AAX26686.
XX
PT Screening for test compounds which inhibit or activate estrogen
PT receptor beta (ER-beta) mediated activation at a API site for use
PT in therapy
XX
PS Claim 14; Page 43-44; 61pp; English.
XX
CC The present sequence represents an estrogen receptor-beta. The
CC specification describes a method for screening test compounds for
CC differential estrogen receptor (ER) alpha-mediated and ER beta-mediated
CC activation at an API site. The method is used for identifying compounds
CC which can be used in therapy.
XX
SQ Sequence 485 AA;

  Query Match   90.4%; Score 2535; DB 20; Length 485;
  Best Local Similarity   99.2%; Pred. No. 7.2e-233;
  Matches 481; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

```
Qy         46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PWCEARSLGHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 120

Qy        166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 180

Qy        226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 240

Qy        286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SAPFTEASMMMLSTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 300

Qy        346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 360

Qy        406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 420

Qy        466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 480

Qy        526 NPQSQ 530
              | |||
Db        481 NLQSQ 485
```

RESULT 11
AAW83212
ID AAW83212 standard; Protein; 477 AA.
XX
AC AAW83212;
XX
DT 20-APR-1998 (first entry)
XX
DE Human oestrogen receptor protein.
XX
KW oestrogen receptor protein; steroid; alternative splicing; estradiol;
KW estone; estriol; screening.
XX
OS Homo sapiens.
XX
PN EP798378-A2.
XX
PD 01-OCT-1997.
XX
PF 25-MAR-1997; 97EP-0200903.
XX
PR 22-NOV-1996; 96EP-0203284.
PR 26-MAR-1996; 96EP-0200820.
XX
PA (ALKO) AKZO NOBEL NV.

```
XX
PI   Dijkema R,  Mosselman S;
XX
DR   WPI; 1997-473188/44.
DR   N-PSDB; AAT88412.
XX
PT   DNA encoding estrogen receptor - useful in screening assay to
PT   identify novel ligands or hormonal analogues
XX
PS   Claim 4; Page 18-20; 45pp; English.
XX
CC   This sequence represents a novel oestrogen binding protein. The cDNA
CC   sequence which encodes this protein can be alternatively spliced
CC   resulting in the detection of additional transcripts (see AAT88413).
CC   This receptor is able to bind and be activated by estradiol, estone and
CC   estriol, can be used in a screening assay for the identification of new
CC   drugs e.g. novel ligands or hormonal analogues.
XX
SQ   Sequence   477 AA;

  Query Match             89.9%;  Score 2522;  DB 18;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-231;
  Matches  477;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          54 MNYSIPSNVTNLEGGPGRQTTSPNVIWPTPGHLSPLVVHRQLSHLYAEPQKSPWCFARSL 113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MNYSIPSNVTNLEGGPGRQTTSPNVIWPTPGHLSPLVVHRQLSHLYAEPQKSPWCFARSL 60

Qy         114 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 120

Qy         174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRKERGGRMLKHKRQR 233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRKERGGRMLKHKRQR 180

Qy         234 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 240

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 300

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 360

Qy         414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 473
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 420

Qy         474 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 477


RESULT 12
AAB47834
ID   AAB47834 standard; Protein; 477 AA.
XX
AC   AAB47834;
XX
DT   25-MAR-2002  (first entry)
XX
DE   Estrogen receptor
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER-alpha; ER-beta; exon 8;
KW   estradiol; nuclear receptor; progesterone receptor
XX
OS   Homo sapiens.
XX
PN   EP1162264-A2.
XX
PD   12-DEC-2001.
XX
PF   25-MAR-1997; 2001EP-0202021.
XX
PR   26-MAR-1996;   96EP-0200820.
PR   22-NOV-1996;   96EP-0203284.
PR   25-MAR-1997;   97EP-0200903.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Mosselman S,  Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   N-PSDB; AAI72128.
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor -
XX
PS   Example A, Page 15-17; 35pp; English.
XX
CC   This sequence shows a novel estrogen receptor (ER). The gene encoding
CC   this new ER is located on chromosome 14 and has a different tissue
CC   distribution from classical ER.  This ER also has two orphan ER's,
CC   ER alpha and ER-beta.  These orphan receptors have estrogen receptor
CC   related structure but do not appear to able to bind estradiol or other
CC   ER ligands. The DNA binding domain (DBD) and ligand binding domain (LBD)
CC   from this ER may be used in the chimeric receptor of the invention
CC   which also has an N terminal domain.  The chimeric receptor, or DNA
CC   encoding it, is useful in a screening assay for identification of new
CC   drugs. Similar chimeric receptors comprising the LBD of the new ER,
CC   and also comprising the DBD and an N-terminal domain derived from
CC   another nuclear receptor e.g., progesterone receptor, are useful for the
CC   screening of compounds to identify new ligands or hormone analogs which
CC   are able to activate the new ER. Chimeric receptors comprising a DBD
CC   of the new ER, and LBD and an N-terminal domain derived from
CC   another nuclear receptor, can be used to identify new ligands or
CC   hormone analogs for the nuclear receptors.
XX
SQ   Sequence   477 AA;

  Query Match             89.9%;  Score 2522;  DB 23;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-231;
  Matches  477;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          54 MNYSIPSNVTNLEGGPGRQTTSPNVIWPTPGHLSPLVVHRQLSHLYAEPQKSPWCFARSL 113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MNYSIPSNVTNLEGGPGRQTTSPNVIWPTPGHLSPLVVHRQLSHLYAEPQKSPWCFARSL 60

Qy         114 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 120

Qy         174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRKERGGRMLKHKRQR 233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRKERGGRMLKHKRQR 180

Qy         234 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 240

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 300

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 360

Qy         414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 473
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 VIATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 420

Qy        474 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 477


RESULT 13
AAW98125
ID   AAW98125 standard; Protein; 549 AA.
XX
AC   AAW98125;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Mouse oestrogen receptor beta isoform mER beta-1.
XX
KW   Oestrogen receptor; ER beta-1; ER beta-c; mouse; tumour;
KW   diagnosis; therapy; isoform.
XX
OS   Mus musculus.
XX
PN   WO9905170-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;   98WO-US15539.
XX
PR   30-JUL-1997;   97US-0054210.
PR   28-JUL-1997;   97US-0053869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Denton R,  Frasion C,  Galien R,  Rische Rigon M;
XX
DR   WPI; 1999-142852/12.
XX
PT   Human oestrogen receptor beta-3 - useful in treating and monitoring
PT   tumours
XX
PS   Example 2; Fig 2a; 89pp; English.
XX
CC   The present sequence is alternatively spliced isoform mER beta 1
CC   of the murine complete oestrogen receptor beta (ER beta-c, or
CC   mER beta 4).  The mER beta 4 gene (see AAX24935) contains 9 exons,
CC   including the newly described exon 5B, which is not found in ER
CC   beta 1.  Exon 5B is located in the ligand binding domain of mER
CC   beta 4, and probably plays a significant role in mER beta-4 function.
CC   The mER beta-1 clone was isolated by RT PCR of mRNA, and by
CC   amplification from a mouse embryonic stem cell genomic DNA library.
CC   It lacks exons 5B.  Preliminary data indicate that the mER beta-1
CC   isoform may be more active than the full-length mER beta-4 (see
CC   AAW98128).  2 other isoforms, mER beta-2 (see AAW98126) and rat rER
CC   beta 4 (see AAW98127), have also been isolated, and the human exon 5B
CC   region (see AAX24934) has been identified.  Based on these sequences,
CC   the invention provides methods for identifying agents that block or
CC   augment ER beta-c mediated transcriptional regulation, methods to
CC   determine whether ER beta-c or its isoforms is being expressed in
CC   tissues or cells, and methods of identifying and using agents that
CC   block the transcriptional regulation of genes by ER beta-c or its
CC   isoforms, which in turn modulates other biological and pathological
CC   processes.  Gene therapy expression of ER beta c is envisaged.
XX
SQ   Sequence   549 AA;

  Query Match             89.2%;  Score 2502;  DB 20;  Length 549;
  Best Local Similarity   88.7%;  Pred. No. 1.2e-229;
  Matches  470;  Conservative   25;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:||||||||| ||:|||||||||||||| |||||||||:| ||| ||||||||||||:||
Db         20 MEIKNSPSSLTSPASYNCSQSILPLEHGPIYIPSSYVESRHEYSAMTFYSPAVMNYSVPS 79

Qy         61 NVTNEIEGCPCRQTTSPNVLWPTFCHELSPLVVHEQLSHLYAEPQKSFWCEAKSLEHTLPVN 120
              :  |||||| ||| |||||||| |||||  | | | ||||||||||||||||||||||
Db         80 STGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEAKSLEHTLPVN 139

Qy        121 METLPPKVSGNPCASPVTCPCSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||: |: |||||| | :|||||||||||||||||||||||||||||||||||||
Db        140 METLKRKLGGSGCASPVTSPSAKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 199

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||:|||||| ||:|
Db        200 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSASEQVH 259

Qy        241 CAGKAKRSGCHAPRVREILLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  |||| || |||:|||||::|||||||||||||||||:||:|||| ||||||||||||
Db        260 CLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLIK 319

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||||||| ||||||||||||||:|||||||||||||||||||
Db        320 LADKELVHMIGWAKKIPGFVELSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 379

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              |||||||||||||||||||||:|||||||||||||||||||||||||||||||| ||:|:|
Db        380 VLDRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEA 439

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :|||| ||||||||||||||:|||||||||:|||||||||||||| |||||||||:||
Db        440 ESSRKLTHLLNAVTDALVWVISKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMK 499

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              |||||||||||||||| ||| ||||:||||   |||||||||| |||
Db        500 CKNVVPVYDLLLEMLNAHTLRGYKSSISGSECCSTEDSKSKEGSQNLQSQ 549


RESULT 14
AAY04434
ID   AAY04434 standard; Protein; 549 AA.
XX
AC   AAY04434;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Murine mER beta-1 clone protein sequence.
XX
KW   Estrogen receptor beta; oestrogen receptor beta; ER beta.
XX
OS   Mus sp.
XX
PN   WO9905171-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;   98WO-US15540.
XX
PR   30-JUL-1997;   97US-0054210.
PR   28-JUL-1997;   97US-0053869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Baron R,  Chambon P,  Denton R,  Dupont S,  Garnier J;
XX
DR   WPI; 1999-142854/12.
XX
PT   Novel murine oestrogen receptor-beta genes and splice variants
PT   useful for treating oestrogen receptor beta mediated disorders
XX
PS   Claim 9; Fig 2a; 89pp; English.
XX
CC   The present invention describes the murine complete oestrogen receptor
CC   beta (mER beta c; also designated mER beta-4). The present sequence
CC   represents mER beta-1, which is an isoform of mER beta c. The protein
CC   can be used to screen for agents that modulate or block the interaction
CC   between the ER beta c protein and physiological, stilbene or
CC   triphenylethylene (anti)oestrogens. The antibody which binds the protein
```

```
CC   can be used in assays to detect activation of the ER beta protein and
CC   can be used for subcellular localisation of the protein. The
CC   complementary nucleic acid can be used to inhibit expression of the
CC   ER beta c nucleic acid.
XX
SQ   Sequence   549 AA;

  Query Match             89.1%;  Score 2500;  DB 20;  Length 549;
  Best Local Similarity   88.5%;  Pred. No. 1.9e-229;
  Matches  469;  Conservative   26;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:||||||||| ||:|||||||||||||| |||||||||:| ||| ||||||||||||:||
Db         20 MEIKNSPSSLTSPASYNCSQSILPLEHGPIYIPSSYVESRHEYSAMTFYSPAVMNYSVPS 79

Qy         61 NVTNLEGGPGRQTTSPNVIWPTPGHISPLVVHPQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              :  |||||| ||| |||||||| |||||   | | | ||||||||||||||||||||||
Db         80 STGNLEGGPVRQTASPNVIWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||: |: |||||| | :|||||||||||||||||||||||||||||||||||||||
Db        140 RETLKRKLGGSGCASPVTSPSAKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 199

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSADFQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||| ||:|
Db        200 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSASEQVH 259

Qy        241 CAGKAKRSGGHAPRVPRLLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  |||| || |||:||||::|||||||||||||||:|:::|||| ||||||||||||
Db        260 CLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVIVSRPSMPFTEASMMMSLTK 319

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||||||| ||||||||||||||:|||||||||||||||||||||
Db        320 LADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 379

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQSA 420
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||| ||:|:|
Db        380 VLDRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEA 439

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :||||| ||||||||||||||:|||||||||:|||||||||||| ||||||||||:||
Db        440 ESSRKITHLINAVTDAIVWVISKSGISSQQQSVRLANLLMLLSHVPHISNKGMEHLLSMK 499

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              |||||||||||||||||| ||| ||||:|||   |||||||||| |||
Db        500 CKNVVPVYDLLLEMLNAHTLRGYKSSISGSECCSTEDSKSKEGSQNLQSQ 549


RESULT 15
AAW98128
ID   AAW98128 standard; Protein; 567 AA
XX
AC   AAW98128;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Mouse complete oestrogen receptor beta mER beta-3.
XX
KW   Oestrogen receptor; ER beta-3; ER beta-c; mouse; tumour;
KW   diagnosis; therapy.
XX
OS   Mus musculus.
XX
PN   WO9905170-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;   98WO-US15539.
XX
PR   30-JUL-1997;   97US-0054210.
PR   28-JUL-1997;   97US-0053869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Denton R, Frasion C, Galien R, Rische-Pigon M;
XX
DR   WPI; 1999-142852/12.
DR   N-PSDB; AAX24935.
XX
PT   Human oestrogen receptor beta-3 - useful in treating and monitoring
PT   tumours
XX
PS   Example 1; Fig 1b; 89pp; English.
XX
CC   The present sequence is the murine complete oestrogen receptor-beta
CC   (ER beta c) mER beta-3.  The mER beta 3 cDNA clone (see AAX24935)
CC   contains 9 exons, including the newly described exon 5B, which is
CC   not found in ER beta-1.  Exon 5B is located in the ligand binding
CC   domain of mER beta-3, and probably plays a significant role in mER
CC   beta-3 function.  The mER beta-3 clone was twice isolated using 2
CC   separate procedures: RT PCR of mRNA; and amplification from a
CC   mouse embryonic stem cell genomic DNA library.  2 Alternatively
CC   spliced forms of mER beta-3 were also identified; mER beta 1 (see
CC   AAW98125), which lacks exon 5B; and mER beta-2 (see AAW98126), which
CC   lacks exon 3.  A third isoform, rER beta-4 (see AAW98127) was
CC   identified in rat ovaries.  The human exon 5B region (see AAX24934)
CC   has also been identified.  Based on these sequences, the invention
CC   provides methods for identifying agents that block or augment
CC   ER beta-c mediated transcriptional regulation, methods to determine
CC   whether ER beta-c or its isoforms is being expressed in tissues or
CC   cells, and methods of identifying and using agents that block the
CC   transcriptional regulation of genes by ER beta-c or its isoforms,
CC   which in turn modulates other biological and pathological
CC   processes.  Gene therapy expression of ER beta-c is envisaged.
XX
SQ   Sequence   567 AA;

  Query Match             88.4%;  Score 2481;  DB 20;  Length 567;
  Best Local Similarity   85.8%;  Pred. No. 1.3e-227;
  Matches  470;  Conservative   25;  Mismatches   35;  Indels   18;  Gaps    1;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:||||||||| ||:|||||||||||||| |||||||||:| ||| ||||||||||||:||
Db         20 MEIKNSPSSLTSPASYNCSQSILPLEHGPIYIPSSYVESRHEYSAMTFYSPAVMNYSVPS 79

Qy         61 NVTNLEGGPGRQTTSPNVIWPTPGHISPLVVHPQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              :  |||||| ||| |||||||| |||||   | | | ||||||||||||||||||||||
Db         80 STGNLEGGPVRQTASPNVIWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||: |: |||||| | :|||||||||||||||||||||||||||||||||||||||
Db        140 RETLKRKLGGSGCASPVTSPSAKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 199

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSADFQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||| ||:|
Db        200 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSASEQVH 259

Qy        241 CAGKAKRSGGHAPRVPRLLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  |||| || |||:||||::|||||||||||||||:|:::|||| ||||||||||||
Db        260 CLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVIVSRPSMPFTEASMMMSLTK 319

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||||||| ||||||||||||||:|||||||||||||||||||||
Db        320 LADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 379

Qy        361 VLD-          -----  -  RDPQKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKA 402
              |||                  ||||||||||||||||||||||:|||||||||||||||
Db        380 VLDRSSEDPHWHVAQTFSAVPPDFQKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKA 439

Qy        403 MILLNSSMYPLVTATQSADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLL 462
              |||||||||||  ||:|:|:|||||| ||||||||||||||:|||||||||:||||||||
Db        440 MILLNSSMYPLATASQEAESSRKLTHLLNAVTDALVWVISKSGISSQQQSVRLANLLMLL 499

Qy        463 SHVPHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGYKSSITGSECSPAEDSKSKE 522
              ||||| |||||||||:|||||||||||||||||||| ||| ||||:|||   |||||||
Db        500 SHVPHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSECCSTEDSKSKE 559
```

```
Qy          523 GSQNPQSQ 540
                ||||  |||
Db          560 GSQNLQSQ 567


RESULT 16
AAY04433
ID   AAY04433 standard; Protein; 560 AA.
XX
AC   AAY04433;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Murine mER beta-3 clone protein sequence.
XX
KW   Estrogen receptor beta; oestrogen receptor beta; ER beta.
XX
OS   Mus sp.
XX
PN   WO9905171-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;   98WO-US15540.
XX
PR   30-JUL-1997;   97US-0054210.
PR   28-JUL-1997;   97US-0053869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Baron R,  Chambon P,  Denton R,  Dupont S,  Garnier J;
XX
DR   WPI; 1999-142853/12.
DR   N-PSDB; AAX83382.
XX
PT   Novel murine oestrogen receptor-beta genes and splice variants -
PT   useful for treating oestrogen receptor-beta mediated disorders
XX
PS   Claim 9; Fig 1b-1; 89pp; English.
XX
CC   The present sequence is the murine complete oestrogen receptor beta
CC   (mER beta c; also designated mER beta 3). The protein can be used to
CC   screen for agents that modulate or block the interaction between the
CC   ER beta c protein and physiological, stilbene or triphenylethylene
CC   (anti)oestrogens. The antibody which binds the protein can be used in
CC   assays to detect activation of the ER beta protein and can be used for
CC   subcellular localisation of the protein. The complementary nucleic acid
CC   can be used to inhibit expression of the ER beta c nucleic acid.
XX
SQ   Sequence   560 AA;

  Query Match             87.6%;  Score 2457;  DB 20;  Length 560;
  Best Local Similarity   85.8%;  Pred. No. 2.5e-225;
  Matches  464;  Conservative   25;  Mismatches   34;  Indels   18;  Gaps    1;

Qy            1 MDIKNSPSSLNSPSSYNCSQSTLPLFHGSLYIPSSYVDSHHEYPAMIPYSPAVMNYSIPS 60
                |:|||||||| ||:|||||||||||||| |||||||||:| ||| ||||||||||||||:||
Db           20 MEIKNSPSSLISPASYNCSQSTLPLFHGPLYIPSSYVESPHEYSAMIPYSPAVMNYSVPS 79

Qy           61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
                :  |||||| ||| |||||||| ||||||  | | | ||||||||||||||||||||||
Db           80 STGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139

Qy          121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
                |||||||: |: |||||| | :|||||||||||||||||||||||||||||||||||||||
Db          140 RETLKRKLGGSGCASPVTSPSAKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 199

Qy          181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRRQRSADEQLH 240
                ||||||||||||||||||||||||||||||||||||||||||||||:|||||| ||:|
Db          200 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYKIVRRQRSASEQVH 259

Qy          241 CAGKAKRSGCHAPRVREILLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
                .                .
Db          260 CLNKAAKISGHFPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTK 319

Qy          301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
                |||||||||| ||||||||||||| ||||||||||||||||:|||||||||||||||||||
Db          320 LADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 379

Qy          361 VLD-------- ---- ----RDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKA 402
                |||                 |||||||||||||||||||||:|||||||||||||||
Db          380 VLDRSSEDPHWHVAQTKSAVPRDEGKCVEGILEIFDMLLATTAREFELKLQHKEYLCVKA 439

Qy          403 MILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLL 462
                |||||||||| ||:|:|:|||| |||||||||||||:|||||||||:|||||||||||||
Db          440 MILLNSSMYPLATASQEAESSRKLTHLLNAVTDALVWVISKSGISSQQQSVRLANLLMLL 499

Qy          463 SHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKE 522
                ||||| |||||||||:|||||||||||||||||||| ||| |||||||||    |||||||
Db          500 SHVRHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSECSSTEDSKSKE 559

Qy          523 G 523
                |
Db          560 G 560


RESULT 17
AAW14723
ID   AAW14723 standard; Protein; 485 AA.
XX
AC   AAW14723;
XX
DT   08-JUN-1997  (first entry)
XX
DE   Rat oestrogen receptor beta (ER-beta).
XX
KW   Orphan receptor; oestrogen receptor beta; ER-beta;
KW   nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW   ovary cancer; cardiovascular disease; osteoporosis;
KW   environment; pollutant.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   Domain          104..169
FT                   /label= DNA-binding_domain
FT   Domain          259..457
FT                   /label= Ligand binding_domain
XX
PN   WO9709348-A2.
XX
PD   14-MAR-1997.
XX
PF   09-SEP-1996;   96WO-EP03933.
XX
PR   08-MAY-1996;   96GB-0009576.
PR   08-SEP-1995;   95GB-0018272.
PR   15-MAR-1996;   96GB-0005550.
PR   11-APR-1996;   96GB-0007532.
XX
PA   (KARO-) KARO BIO AB.
XX
PI   Enmark E,  Gustafsson J,  Kuiper GG;
XX
DR   WPI; 1997-192842/17.
DR   N-PSDB; AAT62842.
XX
PT   New isolated oestrogen receptor beta - used to develop prods. for
PT   treating e.g. cancers, CNS diseases, osteoporosis or cardiovascular
PT   disease
XX
PS   Claim 1; Fig 1; 45pp; English.
XX
CC   A novel rat orphan receptor (AAW14723) is related to the known
CC   oestrogen receptor ER-alpha, and has been designated ER beta.
```

```
CC   It is an oestrogen receptor-related nuclear receptor.  The ER-beta
CC   amino acid sequence was deduced from a cDNA clone (AAI62842) isolated
CC   from a rat prostate cDNA library.  Rat, human and mouse ER-beta
CC   (AAW14723-25) can be used to isolate molecules for use in the
CC   treatment of cardiovascular diseases, central nervous system
CC   diseases, osteoporosis, prostate or ovarian cancer or benign
CC   prostatic hyperplasia and to test environmental chemicals for
CC   oestrogenic activity.
XX
SQ   Sequence   485 AA;

  Query Match             81.7%;  Score 2291;  DB 18;  Length 485;
  Best Local Similarity   88.7%;  Pred. No. 1.4e-209;
  Matches  430;  Conservative   23;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
               ||||||||||||:| : :||:||| | :|||||||| |||||  | | | ||||||||
Db           1 MTFYSPAVMNYSVPGSTSNLDGGPVRLSTSPNVLWPTSGHLSPLATHCQSSLLYAEPQKS 60

Qy         106 PWCEARSLEHTLPVNRETLKRKLSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
               ||||||||||||||||||||||:||: |||||| | :|||||| ||||||||||||||||
Db          61 PWCEARSLEHTLPVNRETLKRKLSGSSCASPVTSPNAKRDAHFCPVCSDYASGYHYGVWS 120

Qy         166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 225
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 180

Qy         226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
               ||:||||||: ||:||  ||||:|||||||:||||  ||||||||||||||||:||:|||
Db         181 YRIVRRQRSSSEQVHCLSKAKRNGGHAPPVKELLISTLSPEQIVLTLLEAPPPNVISRP 240

Qy         286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
               | |||||||||||||||||||||||| ||||||||||||| ||||||||||||||:||||
Db         241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLFDQVRLLESCWMEVLMVGLMW 300

Qy         346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 360

Qy         406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
               |||||||| :| |:|:|||| ||||||||||||||||||||||:||||||||||||||
Db         361 LNSSMYPLASANQEAESSRKLTHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHV 420

Qy         466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEQSQ 525
               || |||||||||:|||||||||||||||||| ||| ||||:|||||  ||||:|| ||
Db         421 RHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSECSSTEDSKNKESSQ 480

Qy         526 NPQSQ 530
               | |||
Db         481 NLQSQ 485


RESULT 18
AAY01596
ID   AAY01596 standard; Protein; 485 AA.
XX
AC   AAY01596;
XX
DT   17-JUN-1999  (first entry)
XX
DE   An estrogen receptor-beta protein.
XX
KW   differential estrogen receptor; ER-alpha; ER-beta; activation;
KW   AP1 site; therapy.
XX
OS   Rattus sp.
XX
PN   WO9911760-A1.
XX
PD   11-MAR-1999.
XX
PF   31-AUG-1998;   98WO-US18030.
XX
PR   04-SEP-1997;   97US-0923708.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Gustafsson J, Kuiper GGJM, Kushner PJ, Nilsson S;
PI   Paech K, Scanlan TS, Webb P;
XX
DR   WPI; 1999-205173/17.
DR   N-PSDB; AAX26685.
XX
PT   Screening for test compounds which inhibit or activate estrogen
PT   receptor beta (ER beta) mediated activation at a AP1 site for use
PT   in therapy
XX
PS   Claim 3; Page 41-43; 61pp; English.
XX
CC   The present sequence represents an estrogen receptor-beta. The
CC   specification describes a method for screening test compounds for
CC   differential estrogen receptor (ER) alpha-mediated and ER beta-mediated
CC   activation at an AP1 site. The method is used for identifying compounds
CC   which can be used in therapy.
XX
SQ   Sequence   485 AA;

  Query Match             81.7%;  Score 2291;  DB 20;  Length 485;
  Best Local Similarity   88.7%;  Pred. No. 1.4e-209;
  Matches  430;  Conservative   23;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
               ||||||||||||:| : :||:||| | :|||||||| |||||  | | | ||||||||
Db           1 MTFYSPAVMNYSVPGSTSNLDGGPVRLSTSPNVLWPISGHLSPLATHCQSSLLYAEPQKS 60

Qy         106 PWCEARSLEHTLPVNRETLKRKLSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
               ||||||||||||||||||||||:||: |||||| | :|||||| ||||||||||||||||
Db          61 PWCEARSLEHTLPVNRETLKRKLSGSSCASPVTSPNAKRDAHFCPVCSDYASGYHYGVWS 120

Qy         166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 225
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 180

Qy         226 YRLVPPQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISPP 285
               ||:||||||: ||:||  ||||:|||||||:||||  ||||||||||||||||:||:|||
Db         181 YRIVRRQRSSSPQVHCLSKAKRNGGHAPRVKELLISTLSPFQIVLTLLEAPPPNVIVSPP 240

Qy         286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
               | |||||||||||||||||||||||| ||||||||||||| ||||||||||||||:||||
Db         241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy         346 RSIDHPGKLIFAPDLVLDRDPGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 360

Qy         406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
               |||||||| :| |:|:|||| ||||||||||||||||||||||:||||||||||||||
Db         361 LNSSMYPLASANQEAESSRKLTHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHV 420

Qy         466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEQSQ 525
               || |||||||||:|||||||||||||||||| ||| ||||:|||||  ||||:|| ||
Db         421 RHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSECSSTEDSKNKESSQ 480

Qy         526 NPQSQ 530
               | |||
Db         481 NLQSQ 485


RESULT 19
AAW14725
ID   AAW14725 standard; Protein; 485 AA.
XX
AC   AAW14725;
XX
DT   08-JUN-1997  (first entry)
XX
```

```
DE   Mouse oestrogen receptor beta (ER-beta).
XX
KW   Orphan receptor; oestrogen receptor beta; ER beta;
KW   nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW   ovary cancer; cardiovascular disease; osteoporosis;
KW   environment; pollutant.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Domain          104..169
FT                   /label= DNA-binding_domain
FT   Domain          260..457
FT                   /label= Ligand-binding_domain
XX
PN   WO9709348-A2.
XX
PD   14-MAR-1997.
XX
PF   09-SEP-1996;    96WO-EP03933.
XX
PR   08-MAY-1996;    96GB-0009576.
PR   08-SEP-1995;    95GB-0018272.
PR   15-MAR-1996;    96GB-0005550.
PR   11-APR-1996;    96GB-0007542.
XX
PA   (KARO ) KARO BIO AB.
XX
PI   Enmark E,  Gustafsson J,  Kuiper GG;
XX
DR   WPI; 1997-192842/17.
DR   N-PSDB; AAT62844.
XX
PT   New isolated oestrogen receptor beta  used to develop prods. for
PT   treating e.g. cancers, CNS diseases, osteoporosis or cardiovascular
PT   disease
XX
PS   Claim 1; Fig 14A; 45pp; English.
XX
CC   A novel mouse orphan receptor (AAW14725) is related to the known
CC   oestrogen receptor ER-alpha, and has been designated ER-beta.
CC   It is an oestrogen receptor-related nuclear receptor.  The ER-beta
CC   amino acid sequence was deduced from a cDNA clone (AAT62844).
CC   Rat, human and mouse ER beta (AAW14723-25) can be used to isolate
CC   molecules for use in the treatment of cardiovascular diseases,
CC   central nervous system diseases, osteoporosis, prostate or ovarian
CC   cancer or benign prostatic hyperplasia and to test environmental
CC   chemicals for oestrogenic activity.
XX
SQ   Sequence   485 AA;

  Query Match             80.8%;  Score 2267;  DB 18;  Length 485;
  Best Local Similarity   88.0%;  Pred. No. 2.7e-207;
  Matches  427;  Conservative   23;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHHQQLSHLYAEPQKS 105
               | ||||||||||:||:  |||||| ||| |||||||| ||||||  | | | ||||||||
Db           1 MAFYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKS 60

Qy         106 PWCEARGLFPTLPVNPETLKRKVSGNPGASPVTGPGGKEDAHFCAVCSDYASGYHYGVWS 165
               |||||||||||||||||||||: |: |||||| | :|||||||||||||||||||||||
Db          61 PWCEARSLEHTLPVNPETLKRKLGGSGGCASPVTSPSTKRDAHFCAVCSDYASGYHYGVWS 120

Qy         166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACRLRKCYEVGMVKCGSRRERCG 225
               |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db         121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACRLRKCYEVGMVKCGSRRERCG 180

Qy         226 YRLVRRQRSADPQLHCACKAFRSGCHAFEVRELLIDALSPEQLVLTLLEAEPPHVLISRT 285
               ||:||||||| ||:||  |||||: || |||:||||::|||||||||||||||:||:|||
Db         181 YRIVRRQRSASEQVHCLNKAKRTSGHTPRVKELLINSLSPEQLVLTLLEAEPPNVLVSRP 240

Qy         286 SAPFTEASMMMSLTKLADKELVHMISWAKKTPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
               | ||||||||||||||||||||||| |||||||||||| ||||||||||||||||:|||| 

Db         241 SMPFTEASMMMSLTKLADKELVHMIGWAKKTPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy         346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
               |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db         301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMIL 360

Qy         406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
               |||||| | ||:|:|:||||| |||||||||||||:|| |||||||:|||||||||||
Db         361 LNSSMYHLATASQEAESSRKLTHLLNAVTDALVWVISKSRISSQQQSVRLANLLMLLSHV 420

Qy         466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
               || |||||||||:|||||||||||||||||||| ||| |||||| |   ||||||||||
Db         421 RHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHILRGYKSSISGSGCGCSIEDSKSKEGSQ 480

Qy         526 NPQSQ 530
               | |||
Db         481 NLQSQ 485


RESULT 20
AAY04435
ID   AAY04435 standard; Protein; 510 AA.
XX
AC   AAY04435;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Murine mER beta-2 clone protein sequence.
XX
KW   Estrogen receptor beta; oestrogen receptor beta; ER beta.
XX
OS   Mus sp.
XX
PN   WO9905171-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;    98WO-US15540.
XX
PR   30-JUL-1997;    97US-0054210.
PR   28-JUL-1997;    97US-0054869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Baron R,  Chambon P,  Denton R,  Dupont S,  Garnier J;
XX
DR   WPI; 1999-142853/12.
XX
PT   Novel murine oestrogen receptor-beta genes and splice variants -
PT   useful for treating oestrogen receptor-beta mediated disorders
XX
PS   Claim 9; Fig 2b; 89pp; English.
XX
CC   The present invention describes the murine complete oestrogen receptor
CC   beta (mER beta c; also designated mER beta-3). The present sequence
CC   represents mER beta-2, which is an isoform of mER beta c. The protein
CC   can be used to screen for agents that modulate or block the interaction
CC   between the ER beta c protein and physiological, stilbene or
CC   triphenylethylene (anti)oestrogens. The antibody which binds the protein
CC   can be used in assays to detect activation of the ER beta protein and
CC   can be used for subcellular localisation of the protein. The
CC   complementary nucleic acid can be used to inhibit expression of the
CC   ER beta c nucleic acid.
XX
SQ   Sequence   510 AA;

  Query Match             80.1%;  Score 2245.5;  DB 20;  Length 510;
  Best Local Similarity   81.3%;  Pred. No. 5.3e-205;
  Matches  441;  Conservative   25;  Mismatches   45;  Indels   39;  Gaps    1;

Qy           1 MDIKNSPSSLNSPSSYNCSQSTLPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
               |:|||||||  ||:|||||||||||||| |||||||||:| ||| ||||||||||||:||
Db          20 MEIKNSPSSLISPASYNCSQSTLPLEHGPIYIPSSYVESHHEYSAMIFYSPAVMNYSVPS 79
```

```
Qy          61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
               :  |||||| ||| |||||||| ||||||  | | | ||||||||||||||||||||||
Db          80 STGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139

Qy         121 RETLKPKVSGNPCASPVTGPGSKRPGAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
               |||||||: |: |||||| | :|||||||||||||||||||||||||||||||||||||
Db         140 RETLKPKIGGSGCASPVISPSAKRPGAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQ-- 197

Qy         181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADFQLH 240
                                                 ||||||||||:|||||| ||:|
Db         198 -----------------------------------GSRRERCGYRIVRRQRSASEQVH 220

Qy         241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
               |  |||| : || |||:|||| ::|||||||||||||||:||:|||| ||||||||||||
Db         221 CLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTK 280

Qy         301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
               |||||||||| |||||||||||| ||||||||||||||:|||||||||||||||||||||
Db         281 LADKELVHMIGWAKKIPGFVELSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 340

Qy         361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
               |||||||||||||||||||||||:|||||||||||||||||||||||||||||||| ||:|:|
Db         341 VLDRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEA 400

Qy         421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
               :|||||  |||||||||||||:|||||||||:|||||||||||||| ||||||||:||
Db         401 ESSRKLTHLLNAVTDALVWVISKSGISSQQQSVRLANLLMLISHVRHISNKGMEHLLSMK 460

Qy         481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECCSTEDSKSKEGSQNPQSQ 530
               |||||||||||||||| ||| ||||:|||| ||||||||||| |||
Db         461 CKNVVPVYDLLLEMLNAHTLRGYKSSISGSECCSTEDSKSKEGSQNLQSQ 510
```

RESULT 21
AAW98126
ID AAW98126 standard; Protein; 510 AA.
XX
AC AAW98126;
XX
DT 21-JUN-1999 (first entry)
XX
DE Mouse oestrogen receptor beta isoform mER beta-2.
XX
KW Oestrogen receptor; ER beta-2; ER beta-c; mouse; tumour;
KW diagnosis; therapy; isoform.
XX
OS Mus musculus.
XX
PN WO9905170-A1.
XX
PD 04-FEB-1999
XX
PF 28-JUL-1998; 98WO US15539.
XX
PR 30-JUL-1997; 97US-0054210.
PR 28-JUL-1997; 97US-0053869.
XX
PA (UYYA) UNIV YALE.
XX
PI Denton R, Frasion C, Galien R, Rische-Rigon M;
XX
DR WPI; 1999-142852/12.
XX
PT Human oestrogen receptor beta-3 - useful in treating and monitoring
PT tumours
XX
PS Example 2; Fig 2b; 89pp; English.
XX
CC The present sequence is alternatively spliced isoform mER beta-2
CC of the murine complete oestrogen receptor-beta (ER beta-c, or
CC mER beta-4). The mER beta-3 gene (see AAX24935) contains 9 exons,
CC including the newly described exon 5B, which is not found in ER
CC beta-1. Exon 5B is located in the ligand binding domain of mER
CC beta-3, and probably plays a significant role in mER beta-3 function.
CC The mER beta-2 clone was isolated from an embryonic stem cell
CC genomic library. It lacks exon 3. 2 other isoforms, mER beta-1
CC (see AAW98125) and rat rER beta-4 (see AAW98127), have also been
CC isolated, and the human exon 5B region (see AAX24934) has been
CC identified. Based on these sequences, the invention provides
CC methods for identifying agents that block or augment ER beta-c
CC mediated transcriptional regulation, methods to determine whether
CC ER beta-c or its isoforms is being expressed in tissues or cells,
CC and methods of identifying and using agents that block the
CC transcriptional regulation of genes by ER beta c or its isoforms,
CC which in turn modulates other biological and pathological
CC processes. Gene therapy expression of ER beta-c is envisaged.
XX
SQ Sequence 510 AA;

```
  Query Match             80.1%;  Score 2245.5;  DB 20;  Length 510;
  Best Local Similarity   81.3%;  Pred. No. 3.3e-205;
  Matches  431;  Conservative   25;  Mismatches   35;  Indels   39;  Gaps    1;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHCSIYIPSSYVESHEEYSAMTPYSPAVMNYSIPS 60
              |:|||||||| ||:|||||||||||||| |||||||:| ||| |||||||||||||:|||
Db         20 MEIKNSPSSLISPASYNCSQSILPLEHGPIYIPSSYVESREEYSAMIPYSPAVMNYSVPS 79

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              :  |||||| ||| |||||||| ||||||  | | | ||||||||||||||||||||||
Db         80 STGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139

Qy        121 RETLKPKVSGNPCASPVTGPGSKRPGAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||: |: |||||| | :|||||||||||||||||||||||||||||||||||||
Db        140 RETLKPKIGGSGCASPVISPSAKRPGAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQ-- 197

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADFQLH 240
                                                ||||||||||:|||||| ||:|
Db        198 ---------------------------------- - -GSRRERCGYRIVRRQRSASEQVH 220

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  |||| : || |||:|||| ::|||||||||||||||:||:|||| ||||||||||||
Db        221 CLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTK 280

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||||||| ||||||||||||||:|||||||||||||||||||||
Db        281 LADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 340

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||||| ||:|:|
Db        341 VLDRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEA 400

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :|||||  |||||||||||||:|||||||||:|||||||||||||| ||||||||:||
Db        401 ESSRKLTHLLNAVTDALVWVISKSGISSQQQSVRLANLLMLISHVRHISNKGMEHLLSMK 460

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECCSTEDSKSKEGSQNPQSQ 530
              |||||||||||||||| ||| ||||:|||| ||||||||||| |||
Db        461 CKNVVPVYDLLLEMLNAHTLRGYKSSISGSECCSTEDSKSKEGSQNLQSQ 510
```

RESULT 22
AAY51946
ID AAY51946 standard; Protein; 474 AA.
XX
AC AAY51946;
XX
DT 23-JUN-2000 (first entry)
XX
DE Bovine ER-beta protein.
XX
KW Bovine; ER-beta; estrogen receptor beta; anticancer anti-osteoporosis;
KW transcriptional transactivator; estrogen-dependent transcription; cancer;
KW estrogen-response element; ovary; prostate; osteoporosis.
XX
OS Bos taurus.

```
XX
PN   DE19849115 A1.
XX
PD   09 MAR 2000.
XX
PF   27 AUG 1998;   98DE-1049115.
XX
PR   27-AUG-1998;   98DE-1049115.
XX
PA   (IHFB ) IHF INST HORMON & FORTPFLANZUNGS.
XX
PI   Walther N;
XX
DR   WPI; 2000-257690/23.
DR   N-PSDB; AAZ89597.
XX
PT   New bovine estrogen-beta receptor, useful for identifying specific
PT   ligands, potentially used for treating cancer and osteoporosis
XX
PS   Claim 1; Page 9-10; 20pp; German.
XX
CC   This invention describes a novel isolated bovine estrogen receptor-beta
CC   (ERb) which has anticancer and anti-osteoporosis activity. ERb is a
CC   transcriptional transactivator that activates estrogen-dependent
CC   transcription of genes that include an estrogen-response element. It is
CC   significantly less active than estrogen receptor-alpha. ERb (also its
CC   isoforms and peptides containing its ligand binding domain), nucleic acid
CC   (III) encoding them, or vectors and cells containing (III) are used to
CC   identify ligands (agonists or antagonists) of ERb, particularly specific
CC   ligands that do not bind to the estrogen receptor-alpha. These ligands
CC   are potentially useful for treating ERb associated disorders,
CC   particularly cancer of the ovary or prostate, or osteoporosis. This
CC   sequence represents the bovine ER-beta protein described in the method of
CC   the invention.
XX
SQ   Sequence   474 AA;

  Query Match             79.6%;  Score 2231.5;  DB 21;  Length 474;
  Best Local Similarity   88.7%;  Pred. No. 6.5e-204;
  Matches  423;  Conservative   17;  Mismatches   34;  Indels    3;  Gaps    1;

Qy          54 MNYSTPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 113
               ||||||    | | |||||||||||||||||||| :| | | |||||||||| | |||
Db           1 MNYSTPD--  NSEDGPGRQTTSPNVLWPTPGHLSPLATHCQPSVLYAEPQKSPWRETRSL 57

Qy         114 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
               |||||||||||||| || |||| | | ||||||||||||||||||||||||||||||||
Db          58 EHTLPVNRETLKRKASGSSCASPATSPSSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 117

Qy         174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 233
               ||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db         118 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 177

Qy         234 SADEQLHCVGKAKRSGGHAPRVPELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
               ::|||||  | ||:||   ||:|||| ||||||||||||||||||||||||||||||||
Db         178 NSDEQLHCLSKTKRNGGPMTRVKELLLSALSPEQLVLTLLEAEPPHVLISRPSTPFTEAS 237

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
               ||||||||||||||||||||||||||||||:|||||||||:||||:|||||||||||||
Db         238 MMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLEVLMVGLMWRSIDHPGK 297

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
               |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db         298 LIFAPDLILDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPS 357

Qy         414 VTATQDADSSKKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 473
               || |:||| ||| |||||||||||||||||:|||||||||||||||||||||||||||||
Db         358 ATAPQEADSGRKLTHLLNAVTDALVWVIAKSGMSSQQQSMRLANLLMLLSHVRHASNKGM 417

Qy         474 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
               |||||||||||||||||||||| ||| || : ||| :  |||:|||||| ||:|
Db         418 EHLLNMKCKNVVPVYDLLLEMLNAHTLRGNKSLVTGSERNLVEDSESKEGSQKPQAQ 474
```

```
RESULT 23
AAW83213
ID   AAW83213 standard; Protein; 416 AA.
XX
AC   AAW83213;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human oestrogen receptor protein exon 8 splice variant.
XX
KW   Oestrogen receptor protein; steroid; alternative splicing; estradiol;
KW   estone; estriol; screening.
XX
OS   Homo sapiens.
XX
PN   EP798378-A2.
XX
PD   01 OCT 1997.
XX
PF   25-MAR-1997;   97EP-0200903.
XX
PR   22-NOV-1996;   96EP-0203284.
PR   25-MAR-1996;   96EP-0200820.
XX
PA   (AKZO ) AKZO NOBEL NV.
XX
PI   Dijkema R,  Mosselman S;
XX
DR   WPI; 1997-473188/44.
DR   N-PSDB; AAT88413.
XX
PT   DNA encoding estrogen receptor - useful in screening assay to
PT   identify novel ligands or hormonal analogues
XX
PS   Claim 4; Page 20-22; 45pp; English.
XX
CC   This partial sequence represents a splice variant of a novel oestrogen
CC   binding protein, detected by screening a human testis cDNA library.
CC   This protein contains an alternative exon 8 (exon 8B) of the ovzel
CC   oestrogen receptor represented. As a consequence of the introduction of
CC   this exon through an alternative splicing reaction, the reading frame
CC   encoding the novel receptor is immediately terminated, creating a
CC   truncation of the carboxy terminus of the novel receptor. This receptor
CC   is able to bind and be activated by estradiol, estone and estriol, can be
CC   used in a screening assay for the identification of new drugs e.g. novel
CC   ligands or hormonal analogues. This variant does not contain an AF 2
CC   region and therefore probably lacks the ability to modulate transcription
CC   of target genes in a ligand dependent fashion.
XX
SQ   Sequence   416 AA;

  Query Match             78.4%;  Score 2198;  DB 18;  Length 416;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-201;
  Matches  415;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          54 MNYSTPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MNYSTPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 60

Qy         114 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 120

Qy         174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 180

Qy         234 SADEQLHCVGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SADEQLHCVGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 240

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 300

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 360

Qy         414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 468
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 415


RESULT 24
AAB47837
ID   AAB47837 standard; Protein; 416 AA.
XX
AC   AAB47837;
XX
DT   25-MAR-2002  (first entry)
XX
DE   ER splice variant ER-beta.
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER alpha; ER beta; exon 8; PCR;
KW   estradiol; nuclear receptor; progesterone receptor; amplify; primer;
KW   polymerase chain reaction; AF-2 region; splice variant.
XX
OS   Homo sapiens.
XX
PN   EP1162264-A2.
XX
PD   12-DEC-2001.
XX
PF   25-MAR-1997; 2001EP-0202021.
XX
PR   26-MAR-1996;   96EP-0200820.
PR   22-NOV-1996;   96EP-0203284.
PR   25-MAR-1997;   97EP-0200903.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Mosselman S,  Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   N-PSDB; AAI72153.
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor -
XX
PS   Example B; Page 17-18; 35pp; English.
XX
CC   The sequences given in AAB47836 and AAB47837 show splice variants
CC   of a novel estrogen receptor (ER). The gene encoding this new ER is
CC   located on chromosome 14 and has a different tissue distribution from
CC   classical ER. This ER also has two orphan ER's, ER-alpha and ER-beta.
CC   These orphan receptors have estrogen receptor related structure but do
CC   not appear to able to bind estradiol or other ER ligands. These
CC   splice variants differ from the parent ER around exon 8. One clone
CC   contains exon 8B through alternative splicing. This causes an immediate
CC   termination of the reading thereby creating a truncation at the carboxy
CC   terminus. The other splice variant contains an alternative exon 8, exon
CC   8C, which encodes two C-terminal amino acids and then has a stop codon.
CC   The splice variant proteins do not contain the AF 2 region and therefore
CC   probably lack the ability to modulate transcription of target genes in
CC   a ligand dependant manner. However they may be able to interfere with
CC   the functioning of the WT classical ER or the novel ER of the invention,
CC   either by heterodimerisation or by occupying estrogen response elements
CC   or by interactions with other transcription factors.
XX
SQ   Sequence   416 AA;

  Query Match             78.4%;  Score 2198;  DB 23;  Length 416;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-201;
  Matches  415;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          54 MNYSIPSNVTNLEGGPGPQTTSPNVLWPTPGHLSPLVVHPQLSHLYAPPQKSPWCPAPSL 113
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MNYSIPSNVTNLEGGPGPQTTSPNVLWPTPGHLSPLVVHPQLSHLYAPPQKSPWCPAPSL 60

Qy         114 EHTLPVNRETLKRKVSGNPCASPVTGPGSKEDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 EHTLPVNRETLKRKVSGNPCASPVTGPGSKEDAHFCAVCSDYASGYHYGVWSCEGCKAFF 120

Qy         174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVRRQR 233
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVRRQR 180

Qy         234 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLIFAEPPHVLISRPSAPFTEAS 240

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 300

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 360

Qy         414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 468
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 415


RESULT 25
AAW33214
ID   AAW33214 standard; Protein; 418 AA.
XX
AC   AAW33214;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human oestrogen receptor protein splice variant (exon 8C).
XX
KW   Oestrogen receptor protein; steroid; alternative splicing; estradiol;
KW   estone; estriol; screening.
XX
OS   Homo sapiens.
XX
PN   EP798378-A2
XX
PD   01-OCT-1997.
XX
PF   25-MAR-1997;   97EP-0200903.
XX
PR   22-NOV-1996;   96EP-0203284.
PR   26-MAR-1996;   96EP-0200820.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Dijkema R,  Mosselman S;
XX
DR   WPI; 1997-473188/44.
DR   N-PSDB; AAT88414.
XX
PT   DNA encoding estrogen receptor - useful in screening assay to
PT   identify novel ligands or hormonal analogues
XX
PS   Claim 4; Page 30-32; 45pp; English.
XX
CC   This sequence represents a splice variant of a novel oestrogen binding
CC   protein isolated from human thymus tissue. This protein contains an
CC   alternative exon 8 (exon 8C) of the novel oestrogen receptor represented
CC   in AAT88412. This novel receptor is able to bind and be activated by
CC   estradiol, estone and estriol, can be used in a screening assay for the
```

```
CC   identification of new drugs e.g. novel ligands or hormonal analogues.
CC   This variant does not contain an AF-2 region and therefore probably
CC   lacks the ability to modulate transcription of target genes in a ligand
CC   dependent fashion.
XX
SQ   Sequence   418 AA;

  Query Match             78.4%;  Score 2198;  DB 18;  Length 418;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-201;
  Matches  415;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          54 MNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 60

Qy         114 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 120

Qy         174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 180

Qy         234 SADEQLHCAGKAKRSGGHAPRVREILLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SADEQLHCAGKAKRSGGHAPRVREILLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 240

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 300

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 360

Qy         414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 468
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 415


RESULT 26
AAB47846
ID   AAB47846 standard; Protein; 418 AA.
XX
AC   AAB47846;
XX
DT   25-MAR 2002  (first entry)
XX
DE   ER splice variant ER alpha.
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER alpha, ER beta, exon 8, PCR,
KW   estradiol; nuclear receptor; progesterone receptor; amplity; primer;
KW   polymerase chain reaction; AF-2 region, splice variant..
XX
OS   Homo sapiens.
XX
PN   EP1162264 A2
XX
PD   12 DEC 2001.
XX
PF   25 MAR 1997; 2001EP-0202021.
XX
PR   26-MAR-1996;   96EP-0200820.
PR   22-NOV-1996;   96EP-0203284.
PR   25-MAR-1997;   97EP-0200903.
XX
PA   (AKZO ) AKZO NOBEL NV.
XX
PI   Mosselman S,  Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   N-PSDB; AAI72146.
```

```
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor
XX
PS   Example 8; Page 23-25; 35pp; English.
XX
CC   The sequences given in AAB47846 and AAB47847 show splice variants
CC   of a novel estrogen receptor (ER). The gene encoding this new ER is
CC   located on chromosome 14 and has a different tissue distribution from
CC   classical ER.  This ER also has two orphan ER's, ER-alpha and ER-beta.
CC   These orphan receptors have estrogen receptor related structure but do
CC   not appear to able to bind estradiol or other ER ligands. These
CC   splice variants differ from the parent ER around exon 8.  One clone
CC   contains exon 8B through alternative splicing.  This causes an immediate
CC   termination of the reading thereby creating a truncation at the carboxy
CC   terminus.  The other splice variant contains an alternative exon 8, exon
CC   8C, which encodes two C-terminal amino acids and then has a stop codon.
CC   The splice variant proteins do not contain the AF-2 region and therefore
CC   probably lack the ability to modulate transcription of target genes in
CC   a ligand dependant manner.  However they may be able to interfere with
CC   the functioning of the WT classical ER or the novel ER of the invention,
CC   either by heterodimerisation or by occupying estrogen response elements
CC   or by interactions with other transcription factors.
XX
SQ   Sequence   418 AA;

  Query Match             78.4%;  Score 2198;  DB 23;  Length 418;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-201;
  Matches  415;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          54 MNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 113
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 60

Qy         114 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 120

Qy         174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 233
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR 180

Qy         234 SADEQLHCAGKAKRSGGHAPRVREILLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SADEQLHCAGKAKRSGGHAPRVREILLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 240

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 300

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 360

Qy         414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 468
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 415


RESULT 27
AAW98127
ID   AAW98127 standard; Protein; 414 AA.
XX
AC   AAW98127;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Rat oestrogen receptor beta isoform mER beta 4.
XX
KW   Oestrogen receptor; ER beta-4; ER beta 4; rat; tumour;
KW   diagnosis; therapy; isoform.
XX
```

```
OS   Rattus norvegicus.
XX
PN   WO9905170-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;   98WO-US15539.
XX
PR   30-JUL-1997;   97US-0054210.
PR   28-JUL-1997;   97US-0054869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Denton R, Frasion C, Callen R, Rische-Rigen M;
XX
DR   WPI; 1999-142852/12.
XX
PT   Human oestrogen receptor beta-3 - useful in treating and monitoring
PT   tumours
XX
PS   Example 2; Fig 2c; 89pp; English.
XX
CC   The present sequence is alternatively spliced isoform rER beta-4
CC   of the complete oestrogen receptor-beta (ER beta-c). Murine ER
CC   beta-c clone mER beta-3 (see AAX24935) has been identified, which
CC   includes newly described exon 5B, which is located in the ligand
CC   binding domain of the receptor and is probably involved in receptor
CC   function. rER beta-4 nucleic acid was obtained from rat ovaries.
CC   It includes exon 5B but lacks exon 6, causing a frameshift and
CC   truncation of the protein 13 amino acids beyond the translated
CC   exon 5B. Also disclosed are mER beta-1 (AAW98125) and mER beta-2
CC   (AAW98126) isoforms, and human exon 5B (AAV24934). Based on these
CC   sequences, the invention provides methods for identifying agents
CC   that block or augment ER beta c mediated transcriptional regulation,
CC   methods to determine whether ER beta-c or its isoforms is being
CC   expressed in tissues or cells, and methods of identifying and using
CC   agents that block the transcriptional regulation of genes by ER
CC   beta-c or its isoforms, which in turn modulates other biological
CC   and pathological processes. Gene therapy expression of ER beta-c
CC   is envisaged.
XX
SQ   Sequence   414 AA;

  Query Match             62.2%;  Score 1746;  DB 20;  Length 414;
  Best Local Similarity   87.6%;  Pred. No. 1.1e-157;
  Matches  319;  Conservative   20;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:|||||||:||:|||||||||||||| |||||||||: ||| ||||||||||||:|
Db         20 MEIKNSPSSLSSPASYNCSQSILPLEHGPIYIPSSYVDNRHEYSAMTFYSPAVMNYSVPG 79

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              : :|||:||| | :|||||||| |||||  | | | ||||||||||||||||||||||
Db         80 STSNLDGGPVRLSTSPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||:||: |||||| | :|||||| ||||||||||||||||||||||||||||||
Db        140 RETLKRKLSGSSCASPVTSPNAKRDAHFCPVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 199

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||: ||:|
Db        200 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSSSEQVH 259

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  ||||:|||||||:|||| ||||||||||||||||:||:|||| ||||||||||||
Db        260 CLNKAKRNGGHAPRVKELLLLSTLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTK 319

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||||||| ||||||||||||||:|||||||||||||||||||
Db        320 LADKELVHMIGWAKKIPGFVELSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 379

Qy        361 VLDR 364
              ||||
Db        380 VLDR 383


RESULT 28
AAY04436
ID   AAY04436 standard; Protein; 730 AA.
XX
AC   AAY04436;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Rat rER beta-4 clone protein sequence.
XX
KW   Estrogen receptor beta, oestrogen receptor beta, ER beta.
XX
OS   Rattus sp.
XX
PN   WO9905171-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;   98WO-US15540.
XX
PR   30-JUL-1997;   97US-0054210.
PR   28-JUL-1997;   97US-0054869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Baron R, Chambon P, Denton R, Dupont S, Garnier J;
XX
DR   WPI; 1999-142853/12.
XX
PT   Novel murine oestrogen receptor beta genes and splice variants -
PT   useful for treating oestrogen receptor-beta mediated disorders
XX
PS   Claim 9; Fig 2c; 89pp; English.
XX
CC   The present invention describes the murine complete oestrogen receptor
CC   beta (mER beta c; also designated mER beta-3). The present sequence
CC   represents the rat isoform rER beta-4 of mER beta c. The protein can
CC   be used to screen for agents that modulate or block the interaction
CC   between the ER beta c protein and physiological, stilbene or
CC   triphenylethylene (anti)oestrogens. The antibody which binds the protein
CC   can be used in assays to detect activation of the ER beta protein and
CC   can be used for subcellular localisation of the protein. The
CC   complementary nucleic acid can be used to inhibit expression of the
CC   ER beta c nucleic acid.
XX
SQ   Sequence   730 AA;

  Query Match             62.2%;  Score 1746;  DB 20;  Length 730;
  Best Local Similarity   87.6%;  Pred. No. 2.7e-157;
  Matches  319;  Conservative   20;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:|||||||:||:|||||||||||||| |||||||||: ||| ||||||||||||:|
Db         20 MEIKNSPSSLSSPASYNCSQSILPLEHGPIYIPSSYVDNRHEYSAMTFYSPAVMNYSVPG 79

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              : :|||:||| | :|||||||| |||||  | | | ||||||||||||||||||||||
Db         80 STSNLDGGPVRLSISPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||:||: |||||| | :|||||| ||||||||||||||||||||||||||||||
Db        140 RETLKRKISGSSCASPVISPNAKRDAHFCPVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 199

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||: ||:|
Db        200 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSSSEQVH 259

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  ||||:|||||||:|||| ||||||||||||||||:||:|||| ||||||||||||
Db        260 CLSKAKRNGGHAPRVKELLLLSTLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTK 319
```

```
Qy          301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
                ||||||||||| ||||||||||||| |||||||||||||||:||||||||||||||||||||
Db          320 LADKELVHMICWAKEIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 379

Qy          361 VLDR 364
                ||||
Db          380 VLDR 383


RESULT 29
ABB83791
ID   ABB83791 standard; Protein; 581 AA.
XX
AC   ABB83791;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Chemidophorus uniparens oestrogen receptor SEQ ID NO 1.
XX
KW   Chemidophorus uniparens; oestrogen; receptor.
XX
OS   Chemidophorus uniparens.
XX
PN   WO200252010-A1.
XX
PD   04-JUL-2002.
XX
PF   15-NOV-2001; 2001WO-JP09995.
XX
PR   25-DEC-2000; 2000JP-0392262.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Sumida K;
XX
DR   WPI; 2002-500766/53.
DR   N-PSDB; ABN81366.
XX
PT   Chemidophorus uniparens originated estrogen receptor genes, applicable
PT   in test systems for evaluating ability of chemicals to control estrogen
PT   receptor activity, e.g. assay safety of chemicals and detecting
PT   endocrine disrupters
XX
PS   Claim 1; Page 59-62; 81pp; Japanese.
XX
CC   The invention relates to an oestrogen receptor gene (ABN81366) encoding a
CC   581 residue amino acid sequence (ABB83791) or its variant or a sequence
CC   encoded by a polynucleotide sequence having not less than 85 % homology
CC   with (ABN81366) and having oestrogen receptor activity. The genes are
CC   applicable in test systems for evaluating the ability of chemicals to
CC   control the oestrogen receptor activity, e.g. assay safety of chemicals
CC   and detecting endocrine disrupters in environment.
XX
SQ   Sequence   581 AA;

  Query Match             44.0%;  Score 1235.5;  DB 24;  Length 581;
  Best Local Similarity   49.1%;  Pred. No. 1e-108;
  Matches  255;  Conservative   78;  Mismatches  122;  Indels   75;  Gaps   14;

Qy           10 LNSPSSYNCSQSILPLEHGSIYIPSSYVDSH    HEYP -----AMTFYSPAVMNY 56
                ||| |      | :|||     |   ||||:      ||          |   || | :::|
Db           26 LNRP---QLKTPLERP--ISEMYVDSNKTGVFNYPEGATYDFSIAAPVYSSASLSY 76

Qy           57 STPSNVTNLEGGPG          KQTTSPNVLWPTPGHLSFLVVHSQLS   HLYAEPAK 104
                :    ||   | |                || |   |   ||| : |       :|  ||
Db           77 A---STNESFGSGNLGGLHSLNNVPPSPVVFLQTAPQLSPFTHHHNQQVPYYLENEPSS 132

Qy          105 SPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGP--GSKRDAHFCAVCSDYASGYHY 161
                |   ||       | :    : :  | | ::   |     | ::  :||||:|||||||
Db          133 SAMREAPPTAFYRPQS--  ENRHHGGRASNSFKGSLSMESTKFTRYCAVCNDYASGYHY 188

Qy          162 GVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGSRR 221
                ||||||||||||||||||||||||||||||||||||||||||||||||||||:| | |:
Db          189 GVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGLRR 248

Qy          222 ERGSYRLVRRQGSAEQLHCAGKAKRSGGHAPKVRE-----    LLLD      --A 262
                :| | |::: :|  ||           | :|  | |              |::        :
Db          249 DRRGGRMLKHKRQRDE---------LDGRNAVAVTEARNTLWPSPLMIKHSKKNSPALS 299

Qy          263 LSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVE 321
                |: ||:|  ||:|||| |     ||:||:|||:|  || |||:||||||:|||::|||||:
Db          300 LTAEQMVSALLDAEPPIVYSEYDPSKPFSEASVMTLLTNLADRELVHMITWAKRVPGFVD 359

Qy          322 LSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLL 381
                |:| ||| |||  |:|:|||:|||:|||||:|||:|:||| :| |||| :|||||||
Db          360 LALHDQVHLLECAWLEILMIGLIWRSLEHPGKLLFAPNLLLDRSQGMCVEGFVEIFDMLL 419

Qy          382 ATTSRFRFLKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALVWV 440
                ||:|||| : :| :|::|:|::|||||| :|  ::: :  :    :  :|: :|| |  :
Db          420 ATSSRFRMMNLQGEEFVCLKSIILLNSGIYTFLSSTLKSLEEKEHIHRVLDKITDTLIHL 479

Qy          441 IAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
                :||||:| |||  ||| ||::|||:|| |||||||| ||||||||||:|||||||:|| |
Db          480 MAKSGLSLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEMLDAHRL 539


RESULT 30
AAG84509
ID   AAG84509 standard; Protein; 595 AA.
XX
AC   AAG84509;
XX
DT   13-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein mutant M396V.
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator-activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO200142307-A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08554.
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000; 2000JP-0207011.
PR   21-JUL-2000; 2000JP-0220508.
PR   02-AUG-2000; 2000JP-0234054.
PR   03-AUG-2000; 2000JP-0235460.
PR   03-AUG-2000; 2000JP-0235461.
PR   03-AUG-2000; 2000JP-0235464.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-367866/38.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer -
XX
PS   Claim 18; Page 181-184; 276pp; English.
XX
CC   The prosent invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
```

```
CC   peroxisome proliferator-activated receptor protein (PPAR), progesterone
CC   receptor protein (PR), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. These proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             44.0%;  Score 1235.5;  DB 22;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1e-108;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHMQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNPCASPVTGPG-----SKRDAHFCAVCSD 154
               : | :   ||       |    :  |:  |     |     | ---  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGREERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              :||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRERGGYRLVPRQPSAIEQLHCAGKAKRSGGHAPPVP----------------EL 258
              :| | |::| | |::: :|  |:     |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:|||:|||:|::|||||:|||:|:|||::|||||||||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT QDADSSRKLAHLLNAVTD 435
              |||||||||||| : || :|:|:|:|:|||||  :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
               |: ::|:|:|:: ||| |||  ||:|||:|:|| |||||||| :|||||||:|||||||
Db        485 TLIHILMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :||| |
Db        545 DAHRL 549


RESULT 31
AAY21626
ID   AAY21626 standard; protein; 595 AA.
XX
AC   AAY21626;
XX
DT   11-AUG-1999  (first entry)
XX
DE   Ligand binding domain of nuclear receptor hER.
XX
KW   Thyroid hormone receptor; aromatic compound; ligand binding domain;
KW   alpha-glcerophosphate dehydrogenase; cardiac; obesity; triglyceride;
KW   plasma cholesterol; anti-hypertriglyceridaemic; atherosclerosis; GPDH;
KW   thyroid hormone replacement therapy; nuclear receptor.
XX
OS   Homo sapiens.
XX
PN   WO9926966-A2.
XX
PD   03-JUN-1999.
XX
PF   25-NOV-1998;   98WO-US25296.
XX
PR   26-NOV-1997;   97US-0980115.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Apriletti JW,  Baxter JD,  Fletterick RJ,  Kushner PJ;
PI   Scanlan TS,  Shiau AK,  Wagner RL,  West BL;
XX
DR   WPI, 1999-357810/30.
XX
PT   Modulating activity of a thyroid hormone receptor
XX
PS   Disclosure; Fig 3G-R; 447pp; English.
XX
CC   The invention relates to a method for modulating activity of a thyroid
CC   hormone receptor that comprises administration of an aromatic compound
CC   which fits spatially and preferentially into a thyroid hormone ligand
CC   binding domain. The aromatic compound (of a specified formula) can be
CC   used to increase alpha-glcerophosphate dehydrogenase (GPDH) levels, at
CC   levels which do not significantly modify cardiac GPDH levels and are
CC   indicated in the treatment of obesity. The compound also lower total
CC   plasma cholesterol and triglyceride levels and can be used as anti-
CC   hypertriglyceridaemic agents. The compound may also be used for treating
CC   atherosclerosis and may be indicated in thyroid hormone replacement
CC   therapy in patients with compromised cardiac function. Sequences
CC   AAY21621  636 amino acid sequences of ligand binding domains of several
CC   members of the nuclear receptor supertamily.
XX
SQ   Sequence   595 AA;

  Query Match             44.0%;  Score 1233.5;  DB 20;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1.6e-108;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQFSPWC--EARSLEHTLPVNRETLKRKVSGNPCASPVTGPG-----SKRDAHFCAVCSD 154
               : | :   ||       |    :  |:  |     |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGREERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRERGGYRLVPRQPSAIEQLHCAGKAKRSGGHAPPVP---------- ---EL 258
              :| | |::| | |::: :|  |:     |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:|||:|||:|::|||||:|||:|:|||::|||||||||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT QDADSSRKLAHLLNAVTD 435
```

```
              ||||||||:|||| : || :|:::|:|::|||||| :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: ||| |||  ||| ||::||||:|| |||||||| :||||||||:||||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 32
AAG84505
ID   AAG84505 standard; Protein; 595 AA.
XX
AC   AAG84505;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein.
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator-activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer.
XX
OS   Homo sapiens.
XX
PN   WO200142307-A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08553.
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000; 2000JP-0207011.
PR   21-JUL-2000; 2000JP-0220508.
PR   02-AUG-2000; 2000JP-0234053.
PR   03-AUG-2000; 2000JP-0235463.
PR   03-AUG-2000; 2000JP-0235461.
PR   03-AUG-2000; 2000JP-0235464.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-467866/38.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer -
XX
PS   Claim 7; Page 167-170; 276pp; English.
XX
CC   The present invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
CC   peroxisome proliferator activated receptor protein (PPAR), progesterone
CC   receptor protein (PR), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. These proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             44.0%;  Score 1233.5;  DB 22;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1.6e-108;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNQSQSILPLFH--QSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP----QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-  ------- --------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 Q-PGLPVGPGSPAAAFQSNGLGGPPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
               : | :   ||       |    :  |: |    :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305
              :| | |:::| | |::: :|  |:     |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISK--PSAFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: |||:|||||  || |||:|||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:||:|||:|||||:|||:|:|||::||||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||||| :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: ||| |||  ||| ||::||||:|| |||||||| :||||||||:||||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 33
AAG84506
ID   AAG84506 standard; Protein; 595 AA.
XX
AC   AAG84506;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein mutant K303R.
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator-activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO200142307-A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08553.
```

```
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000;   2000JP-0207011.
PR   21-JUL-2000;   2000JP-0220508.
PR   02-AUG-2000;   2000JP-0234053.
PR   03-AUG-2000;   2000JP-0235460.
PR   03-AUG-2000;   2000JP-0235461.
PR   03-AUG-2000;   2000JP-0235463.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-367866/38.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer -
XX
PS   Claim 15; Page 170-174; 276pp; English.
XX
CC   The present invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
CC   peroxisome proliferator activated receptor protein (PPAR), progesterone
CC   receptor protein (PR), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. These proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             44.0%;  Score 1233.5;  DB 22;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1.6e-108;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |       |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--FAPSLPHTLPVNPPTIKPPVSGNPCASPVTGPG-----SKPDAHPCAVCSD 154
              : | :   ||       |    :  |:  |    :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRFRCGYRLVRFQRSADEQLHCAGKAKRSGGHAPRV----------------REL 258
              :| | |::| | |::: :|  |:     |: :   | |  :                |
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LIPAISPPQIVLTLEAPPHVICSE -FSAFPTEASFMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSIFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              |||:|:| ||| ||| |:|:||:||:|::|||||:|||:|:|||:|||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRK---LNAVID 433
              |||||||:|||| : :| :|:|:|:|:|||| :|  :::| :  :     : :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKS---IILLNSGVYTFLSSTLKSLEEKDHIHRVLDKIID 481

Qy        434 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 493
              :|: :||:|:|:: |||  ||| |||:||||:|| ||||||||:|||||||:||||||
Db        482 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 541

Qy        494 NAHVL 498
              :|| |
Db        542 DAHRL 546


RESULT 34
AAG84513
ID   AAG84513 standard; Protein; 595 AA.
XX
AC   AAG84513;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein mutant S578P.
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator-activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO200142307-A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08553.
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000;   2000JP-0207011.
PR   21-JUL-2000;   2000JP-0220508.
PR   02-AUG-2000;   2000JP-0234053.
PR   03-AUG-2000;   2000JP-0235460.
PR   03-AUG-2000;   2000JP-0235461.
PR   03-AUG-2000;   2000JP-0235463.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-367866/38.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer -
XX
PS   Claim 20; Page 194-197; 276pp; English.
XX
CC   The present invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
CC   peroxisome proliferator-activated receptor protein (PPAR), progesterone
CC   receptor protein (PR), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. These proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
```

```
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             44.0%;  Score 1233.5;  DB 22;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1.6e-108;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSTLPLEH--GSTYTPSS--YVDSHHEYPAMTFYSPAVMNYSTPS 60
              |    || |      | :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP---  - QLKTPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ           ------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                    : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||       |    :  |:  |   :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSRRERCGYRLVRRQRSAEQLHCAGKAKRSGGHAPRVR---------------EL 258
              :| | |::| | (::: :|  |:     |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD--  --GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:||||  || |||:|||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:||:|||::|||||:|||:|:|||:|||||||::||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVID 435
              |||||||:|||| : || :|:::|:|::|||||  :|  :::| :  :     :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
                :|| |
Db        545 DAHRL 549


RESULT 35
ABB76478
ID   ABB76478 standard; Protein; 595 AA.
XX
AC   ABB76478;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human nuclear oestrogen receptor alpha.
XX
KW   oestrogen; receptor; human; transgenic mouse; cytostatic;
KW   antiinflammatory; antidiabetic; endocrine; anorectic; hepatotropic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          180..262
FT                   /note= "DNA binding domain"
FT   Region          263..301
FT                   /note= "D hinge region"
FT   Domain          302..552
FT                   /note= "ligand binding domain"
XX

PN   WO200228175-A2.
XX
PD   11-APR-2002.
XX
PF   28-SEP-2001; 2001WO-IB02246.
XX
PR   04-OCT-2000; 2000FR-0012570.
PR   11-MAY-2001; 2001US-0853033.
XX
PA   (ADER-) ADEREGEM ASSOC DEV RECH EN GENETIQ.
XX
PI   Chambon P, Metzger D;
XX
DR   WPI; 2002-463217/49.
DR   N-PSDB; ABL57497.
XX
PT   A transgenic mouse, useful in screening for medicaments for the
PT   treatment of e.g., diabetes or skin cancers, comprises a fusion protein
PT   between a recombinase Cre, and a modified ligand binding domain of the
PT   nuclear estrogen receptor alpha -
XX
PS   Claim 6; Page 131-132; 149pp; English.
XX
CC   The present sequence is the protein sequence of the human nuclear
CC   oestrogen receptor alpha (1). The invention relates to a non-human
CC   metazoan organism, especially a transgenic mouse, characterised in
CC   that at least one cell comprises: (i) a fusion protein formed from
CC   a recombinase (Cre), a hinge region (preferably human (1) D hinge),
CC   and a modified ligand binding domain (LBD) of a nuclear oestrogen
CC   receptor, especially human (1), its fragment or variant; and (ii)
CC   one or more genes or DNA sequences of interest belonging to the
CC   genome of the organism, into which one or more recognition sites
CC   of the recombinase protein are inserted. The (1) LBD domain is
CC   preferably modified by a G521R, G400V, or M543A/L544A mutation.
CC   The fusion protein has negligible, or even zero, recombinase
CC   activity in the presence of a natural ligand such as oestradiol,
CC   but recombinase activity is induced by a small quantity of a
CC   synthetic ligand that has antioestrogenic activity, e.g. tamoxifen
CC   or 4-hydroxytamoxifen. The metazoan organism or its cells, such
CC   as epidermal cells, hepatocytes or adipocytes, are useful in
CC   carrying out a spatiotemporally controlled site-specific
CC   recombination of a DNA sequence of interest in its natural
CC   chromatin environment. It is also used in screening of medicaments
CC   for pathological conditions associated with an alteration of the
CC   expression and/or function of the DNA sequence of interest, such as
CC   skin cancer, inflammation, diabetes, alopecia, obesity, or in
CC   promoting hepatic regeneration.
XX
SQ   Sequence   595 AA;

  Query Match             44.0%;  Score 1233.5;  DB 23;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1.6e-108;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSTLPLEH--GSTYTPSS--YVDSHHEYPAMTFYSPAVMNYSTPS 60
              |    || |      | :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP-      QLKTPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ---- ------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                    : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||       |    :  |:  |   :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSRRERCGYRLVRRQRSAEQLHCAGKAKRSGGHAPRVR----- --- - - - --EL 258
              :| | |::| | (::: :|  |:     |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-  - --GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305
```

```
Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| ||| |:|:||:||:|||:-||||:|||:|:|||--||||||--||
Db        365 PGFVDLTIHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:|:|:|:|||||  :|  :::| :  :  :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: ||| ||| ||::|||:|| |||||||| :|||||||:|||||||
Db        485 TLIHLMAKAGLTIQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 36
AAG84512
ID   AAG84512 standard; Protein; 595 AA.
XX
AC   AAG84512;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein mutant K531E.
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO200142307-A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08553.
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000; 2000JP-0207011.
PR   21-JUL-2000; 2000JP-0220508.
PR   02-AUG-2000; 2000JP-0234053.
PR   03-AUG-2000; 2000JP-0235460.
PR   03-AUG-2000; 2000JP-0235461.
PR   03-AUG-2000; 2000JP-0235463.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-367866/38.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer -
XX
PS   Disclosure; Page 191-194; 276pp; English.
XX
CC   The present invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
CC   peroxisome proliferator activated receptor protein (PPAR), progesterone
CC   receptor protein (PR), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. These proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             43.9%;  Score 1231.5;  DB 22;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 2.5e-108;
  Matches  263;  Conservative   89;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNQSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |   || |      | :|||   | :|: ||   | :: | | | : | | :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |     | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPINSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EAPSLFHTLPVNRETLKRKVSGNPCASPVTGPG-----SKRDAHFCAVCSD 154
               : | :   ||       |    :  |:  |    :      |     | ::  :|||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 269
              :| | |:|| | |::: :|  |:      |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPIRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              |||:|:|: ||| ||| | | || || |||  ||||:|||:|:|||--||||||--||
Db        365 PGFVDLTIHDQVHLLECAWLEILMIGLVWRSVEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| . || .|..|.|..|||| :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVPHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ..|||:|.. |||  ||| ||::|||:|| |||||||| :|||:||||:|||||||
Db        485 TLIHLMAKAGLTIQQQHQRLAQLLLILSHIPHMSNKGMEHLYSMKCENVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 37
AAG84507
ID   AAG84507 standard; Protein; 595 AA.
XX
AC   AAG84507;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein mutant S309F.
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
```

```
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO200142307 A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08553.
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000; 2000JP-0207011.
PR   21-JUL-2000; 2000JP-0220508.
PR   02-AUG-2000; 2000JP-0234053.
PR   04-AUG-2000; 2000JP-0235460.
PR   04-AUG-2000; 2000JP-0235461.
PR   04-AUG-2000; 2000JP-0235463.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-467866/48.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer
XX
PS   Claim 16; Page 174 177; 276pp; English.
XX
CC   The present invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
CC   peroxisome proliferator-activated receptor protein (PPAR), progesterone
CC   receptor protein (PR), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. Those proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             43.9%;  Score 1239.5;  DB 22;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 3.1e-108;
  Matches  264;  Conservative   87;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNQSQSTLPLEH--GSTYTPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |   || |       |  :|||   | :|: ||   | ::  |  |  | : |  |  :
Db         21 NELEPLNRP     --QLKTPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ--          ----------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                    : || :|   |  ||| |  | |    | |
Db         75 Q TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 134

Qy        102 PQKSPWQ--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
                : | :   ||       |    :  |:  |    :      |      | ::  :||||:|
Db        135 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRERGGYRLVRRQRSADEQLHCVRGKAKRSGGHAPEVR--------------EL 258
              :| | |::| | |::: :|  |:     |: :   | |  :|

Db        251 MKGGIRKDRRGGRMLKHKRQRDD---  --GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |   |: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALPLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:|  ||| |||  |:|:||:||:|||::|||||:|:|||::|||||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||| :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 38
AAG84508
ID   AAG84508 standard; Protein; 595 AA.
XX
AC   AAG84508;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein mutant G390D.
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator-activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO200142307 A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08553.
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000; 2000JP-0207011.
PR   21-JUL-2000; 2000JP-0220508.
PR   02-AUG-2000; 2000JP-0234053.
PR   04-AUG-2000; 2000JP-0235460.
PR   04-AUG-2000; 2000JP-0235461.
PR   04-AUG-2000; 2000JP-0235463.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-467866/48.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer -
XX
PS   Claim 17; Page 177 180; 276pp; English.
XX
CC   The present invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
CC   peroxisome proliferator-activated receptor protein (PPAR), progesterone
```

```
CC   receptor protein (ER), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. These proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             43.7%;  Score 1226.5;  DB 22;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 7.5e-108;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSTLPIFH--GSIYIPSS--YVDSHHFYPAMTFYSPAVMNYSIPS 60
              |   || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP  ----QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                : || :|   |  ||| |  | |    | |
Db         75 Q-IGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EAPSLEHTLPVNRETLKPKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
               : |  :   ||       |    :  |:  |    :      |     | ::  :|||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSPRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR----------------EL 258
              :| | |::| | |::: :|  |:    |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||::|::
Db        306 LADSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||: |:||||:|||||.|||.|.|||..||||||..||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIDLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||||  :|  :::| :  :    :  :|: :||
Db        425 FDMILATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::|||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||
Db        485 TLIHIMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 39
AAG84514
ID   AAG84514 standard; Protein; 595 AA.
XX
AC   AAG84514;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human oestrogen receptor alpha protein mutant G390D/S578P
XX
KW   Ligand dependent transcriptional factor; oestrogen receptor; ER;
KW   glucocorticoid receptor protein; GR; mineralocorticoid receptor protein;
KW   MR; peroxisome proliferator activated receptor protein; PPAR;
KW   progesterone receptor protein; PR; pregnane X receptor protein; PXR;
KW   thyroid hormone receptor protein; TR; vitamin D receptor protein; VDR;
KW   transactivation; ERalpha; breast cancer; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO200142307-A1.
XX
PD   14-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-JP08553.
XX
PR   07-DEC-1999;   99JP-0348022.
PR   27-DEC-1999;   99JP-0370667.
PR   07-JUL-2000; 2000JP-0207011.
PR   21-JUL-2000; 2000JP-0220508.
PR   02-AUG-2000; 2000JP-0234053.
PR   03-AUG-2000; 2000JP-0235460.
PR   03-AUG-2000; 2000JP-0235461.
PR   03-AUG-2000; 2000JP-0235463.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Saito K,  Ohe N,  Satoh H;
XX
DR   WPI; 2001-367866/38.
XX
PT   Ligand dependent transcriptional factors, nucleic acids encoding them
PT   and cells comprising them and a specified reporter gene, useful for
PT   screening agents for the treatment of breast cancer -
XX
PS   Claim 21; Page 197-209; 276pp; English.
XX
CC   The present invention relates to ligand dependent transcriptional factors
CC   including oestrogen receptor (ER) alpha and beta protein, glucocorticoid
CC   receptor protein (GR), mineralocorticoid receptor protein (MR),
CC   peroxisome proliferator-activated receptor protein (PPAR), progesterone
CC   receptor protein (PR), pregnane X receptor protein (PXR), thyroid hormone
CC   receptor protein (TR) and vitamin D receptor protein (VDR), the nucleic
CC   acids encoding them and cells comprising them and a specified reporter
CC   gene for the ligand dependent transcriptional factor. These proteins are
CC   useful in the modulation of ligand dependent transcriptional factor
CC   activity. The cells, mutant ERalpha and the polynucleotide encoding it
CC   may be used in assays for qualitatively analysing an activity for
CC   transactivation of a reporter gene by a test ERalpha, for screening
CC   mutant ligand dependent transcriptional factors, for evaluating an
CC   activity for transactivation of a reporter gene by a test ERalpha and/or
CC   for screening a compound useful for treating a disorder of a mutant
CC   ERalpha, especially breast cancer.
XX
SQ   Sequence   595 AA;

  Query Match             43.7%;  Score 1226.5;  DB 22;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 7.5e-108;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSTLPIFH--GSIYIPSS--YVDSHHFYPAMTFYSPAVMNYSIPS 60
              |   || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                : || :|   |  ||| |  | |    | |
Db         75 Q-IGLPYGPGSEAAAPGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EAPSLEHTLPVNRETLKPKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
               : |  :   ||       |    :  |:  |    :      |     | ::  :|||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSPRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR----------------EL 258
```

```
                :| | |:::| | |::: :|  |:     |: :   | |  :|
Db          251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy          259 LLDALSPEQLVILTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
                |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||:: 
Db          306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy          317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
                |||||:|:| ||| ||| |:|:||: |:|||::|||||:|||:|:|:|::|||||||::|| 
Db          365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy          377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
                |||||||:|||| : || :|:::|:|:::|||||:|  :::| :  :    :  :|: :|| 
Db          425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy          436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                 |: ::||:|:: ||  ||| ||:::|||:|| |||||||  :|||||||||:|||||||| 
Db          485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy          496 NAHVL 500
                :|| |
Db          545 DAHRL 549


RESULT 40
AAY83506
ID   AAY83506 standard; Protein; 595 AA.
XX
AC   AAY83506;
XX
DT   19-JAN-2000  (first entry)
XX
DE   Human estrogen receptor protein.
XX
KW   Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KW   huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
KW   atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW   Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW   dentatorubropallidoluysian atrophy; cell proliferation; cell survival;
KW   neoplastic; malignant; autoimmune; fibrotic; estrogen receptor.
XX
OS   Homo sapiens.
XX
PN   WO9945944 A1.
XX
PD   16 SEP 1999.
XX
PF   11 MAR 1999;   99WO US05250.
XX
PR   12 MAR 1998;   98US-0041866.
XX
PA   (BURN ) BURNHAM INST.
XX
PI   Bredesen DE;  Rabizadeh S;
XX
DR   WPI; 1999-561617/47.
DR   N-PSDB; AAZ23433.
XX
PT   New proapoptotic dependence peptides, used to develop products for
PT   treating, e.g. Alzheimer's disease -
XX
PS   Disclosure; Page 185-187; 199pp; English.
XX
CC   This invention describes novel pure proapoptotic dependence peptides
CC   which comprise a sequence of an active dependence domain selected from
CC   dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
CC   huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
CC   SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
CC   of inducing cell death and can be used to develop products to mediate or
CC   inhibit apoptosis. The methods can be used for reducing the severity of
CC   a proapoptotic dependence domain mediated pathological conditions e.g.
CC   Huntington's disease, Alzheimer's disease, Kennedy's disease,
CC   Spinocerebellar ataxias, dentatorubropallidoluysian atrophy,
CC   Machado-Joseph disease, stroke or head trauma. They can also be used for
CC   reducing the severity of a pathological condition mediated by upregulated
CC   cell proliferation or cell survival e.g. neoplastic, malignant,
CC   auto-immune or fibrotic conditions. This sequence represents a human
CC   estrogen receptor described in the method of the invention.
XX
SQ   Sequence   595 AA;

  Query Match          48.7%;  Score 1224.5;  DB 20;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 1.2e-107;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy            5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
                |    || |      | :|||   | :|: ||   | :: |  |  | : |  |  :
Db           21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy           61 NVTNLEGGPGRQ--------- --------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                  | |  ||| :                 : || :|    |  ||| |  | |    | |
Db           75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy          102 PQKSPWC--EAPSLEHILPVNPETLKPKVSGNPCASPVTGPG-----SKRDAHFCAVCSD 154
                 : | :   ||       |    :  |: |   :     |     | ::  :|||||| 
Db          134 NEPSGYIVREAGPPAFYRP---NSDNRRQGGKERLASTNDKGSMAMESAKETRYCAVCND 190

Qy          155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
                |||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||| 
Db          191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy          215 VKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 269
                :| | |:::| | |::: :|  |:     |: :   | |  :|
Db          251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy          270 LLDALSPEQLVILTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 327
                |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||:: 
Db          306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy          328 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 387
                |||||:|:| ||| ||| |:|:||:||:|||::|||||:|||:|:|:|::|||||||::|| 
Db          365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGECVEGMVEI 424

Qy          388 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 446
                |||||||:|||| : || :|:::|:|:::|||||:|  :::| :  :    :  :|: :|| 
Db          425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy          447 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 506
                 |: ::||:|:: |||  ||| ||:::|||:|| |||||||| :|||||||||:|||||||| 
Db          485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy          507 NAHVL 511
                :|| |
Db          545 DAHRL 549


Search completed: June 16, 2003, 13:12:50
Job time : 141.67 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 23:28:27 ; Search time 2888.6 Seconds
(without alignments)
10641.499 Million cell updates/sec

Title: US-08-826-361A-24
Perfect score: 1898
Sequence: 1 CACGAATCTTTGAGAACATT ..TGAATCGAGCGTTTCCACAC 1898

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: qb_est1:*
10: qb_est2:*
11: qb_htc:*
12: qb_est3:*
13: qb_est4:*
14: qb_est5:*
15: em_estfun:*
16: em_estom:*
17: qb_qss:*
18: em_qss_hum:*
19: em_qss_inv:*
20: em_qss_pln:*
21: em_qss_vrt:*
22: em_qss_fun:*
23: em_qss_mam:*
24: em_qss_mus:*
25: em_qss_other:*
26: em_qss_pro:*
27: em_qss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	486.4	25.6	727	12	BG722442	BG722442 602693671
c 2	208	11.0	627	14	BM942391	BM942391 UI-M-BH3
3	188.2	9.9	652	10	BB666338	BB666338 BB666338
4	177.2	9.3	666	10	BB609187	BB609187 BB609187
5	176.4	9.3	627	10	BB666608	BB666608 BB666608
c 6	162.4	8.6	528	14	BM933826	BM933826 UI-M-BH3
7	162.2	8.5	580	12	BG738138	BG738138 fp04t05.y
c 8	159	8.4	531	9	AA224131	AA224131 zr14e02.r
9	157	8.4	642	12	BF058527	BF058527 7k41h08.x
10	157	8.3	782	10	AW134052	AW134052 fi14h07.y
11	155.4	8.2	641	10	AW976636	AW976636 EST388745
12	153.8	8.1	648	12	BF660250	BF660250 maa95g12.
13	149.6	7.9	642	12	BF733032	BF733032 nac18e05.
14	131.4	6.9	594	12	BG084311	BG084311 H3099A08-
15	129.8	6.8	1142	14	BM810012	BM810012 AGENCOURT
16	126.4	6.7	137	17	AF455798	AF455798 AF455798
c 17	121	6.4	558	12	BG891892	BG891892 fp46e11.x
18	117	6.2	692	13	BJ068905	BJ068905 BJ068905
19	115.6	6.1	532	17	AZ596395	AZ596395 1M0409D10
20	115.2	6.1	611	17	AQ719425	AQ719425 HS_5512_B
21	110.6	5.8	1443	11	BC008218	BC008218 Homo sapi
22	108.8	5.7	610	13	BJ069238	BJ069238 BJ069238
23	107.8	5.7	703	12	BG744031	BG744031 602722811
c 24	107.4	5.7	575	12	BF510841	BF510841 UI-H-BI4-
25	107.4	5.7	2199	11	AK619222	AK619222 Mus muscu
26	106.2	5.6	676	12	BG747925	BG747925 602705778
27	102	5.4	844	13	BI100868	BI100868 602886184
28	101.4	5.3	937	13	BM562894	BM562894 AGENCOURT
29	100.4	5.3	558	13	BJ058831	BJ058831 BJ058831
30	99.8	5.3	908	10	BE546976	BE546976 601071950
31	99.6	5.2	800	9	AU169990	AU169990 AU169990
32	99.6	5.2	1004	9	AL527012	AL527012 AL527012
33	96.6	5.1	775	12	BG480960	BG480960 602529118
34	96.4	5.1	979	12	BG033384	BG033384 602301556
35	95	5.0	818	13	BI738331	BI738331 603361587
36	93	4.9	402	12	BF717864	BF717864 fd52h03.y
37	91.8	4.8	669	14	BM957249	BM957249 fy75h10.y
38	91.2	4.8	434	10	BB711891	BB711891 BB711891
39	91.2	4.8	521	12	BG687711	BG687711 334912 BA
40	91.2	4.8	698	10	BB196002	BB196002 BB196002
41	91.2	4.8	1638	11	BC016462	BC016462 Mus muscu
42	91	4.8	399	12	BG010872	BG010872 RC5-GN019
43	90.6	4.8	592	12	BG746447	BG746447 602703728
c 44	90.2	4.8	413	10	AW113877	AW113877 MC809 mou
45	90.2	4.8	1010	9	AL547737	AL547737 AL547737
46	89.6	4.7	275	10	BB196485	BB196485 BB196485
47	89.6	4.7	425	10	BB710863	BB710863 BB710863
48	89.2	4.7	1120	14	BM908889	BM908889 AGENCOURT
49	88.8	4.7	729	13	BG914551	BG914551 602813303
50	88.2	4.6	462	9	AA411285	AA411285 zv26g03.r
c 51	88.2	4.6	477	9	AI341113	AI341113 qx84c02.x
c 52	88.2	4.6	479	9	AI653325	AI653325 wb24c06.x
53	88.2	4.6	549	9	AL047238	AL047238 DKFZp586N
54	88.2	4.6	558	14	BM786552	BM786552 K-EST0065
c 55	88.2	4.6	644	12	BF223014	BF223014 7q27e02.x
56	88.2	4.6	754	12	BG682389	BG682389 602630127
57	88.2	4.6	889	9	AL561541	AL561541 AL561541
58	88.2	4.6	891	9	AL525589	AL525589 AL525589
59	88.2	4.6	909	14	BQ898227	BQ898227 AGENCOURT
60	88.2	4.6	926	9	AL513888	AL513888 AL513888
61	88.2	4.6	960	14	BQ879017	BQ879017 AGENCOURT
62	88.2	4.6	967	9	AL555839	AL555839 AL555839
63	88.2	4.6	972	14	BQ882678	BQ882678 AGENCOURT
64	88.2	4.6	1019	13	BM557938	BM557938 AGENCOURT
65	88.2	4.6	1090	12	BG176463	BG176463 602338137
66	86.8	4.6	674	14	BM791802	BM791802 K-EST0071
67	86.8	4.6	911	14	BQ948248	BQ948248 AGENCOURT
68	86.6	4.6	1176	12	BF162189	BF162189 601770663
69	86.4	4.6	385	10	AV647429	AV647429 AV647429
70	86.4	4.6	399	10	AV647483	AV647483 AV647483
71	86.4	4.6	399	10	AV647486	AV647486 AV647486
c 72	86.4	4.6	446	9	AI636264	AI636264 tz93e10.x
c 73	86.4	4.6	465	9	AI393892	AI393892 tq05e05.x
c 74	86.4	4.6	519	9	AU143800	AU143800 AU143800
75	86.4	4.6	616	12	BG723061	BG723061 602699612
76	86.2	4.5	863	14	BQ949290	BQ949290 AGENCOURT
77	86	4.5	416	13	BI133725	BI133725 UI-M-BH3-
78	86	4.5	794	12	BF138044	BF138044 601784742
c 79	85.8	4.5	271	13	BI300602	BI300602 UI-R-CV2-

	80	85.8	4.5	450	17	AZ042574 AZ042574 RPCI 23 9
	81	85.8	4.5	992	9	AL547687 AL547687
	82	85.8	4.5	1071	14	BM906677 BM906677 AGENCOURT
	83	85.8	4.5	4141	11	BC030234 BC030234 Homo sapi
c	84	85.6	4.5	279	12	BC374797 BC374797 UI-R-CV1
c	85	85.6	4.5	279	13	BI298646 BI298646 UI-R-CV2
	86	85.6	4.5	569	10	AV968271 AV968271 AV968271
c	87	85.4	4.5	758	14	BQ446111 BQ446111 UI-H-EU1
	88	85.2	4.5	645	10	BE533846 BE533846 601233424
	89	85.2	4.5	937	9	AL527466 AL527466 AL527466
	90	85	4.5	910	9	AL556302 AL556302 AL556302
	91	85	4.5	2468	11	AK004709 AK004709 Mus musc
c	92	84.8	4.5	345	12	BP551955 BP551955 UI-R-C2p
c	93	84.8	4.5	384	9	AA851419 AA851419 EST194087
c	94	84.8	4.5	384	9	AA893618 AA893618 EST197421
c	95	84.8	4.5	390	14	W13715 W13715 ma83b04.r1
	96	84.8	4.5	428	9	AI323943 AI323943 ma83b04.x
c	97	84.8	4.5	451	9	AI327473 AI327473 ma83b04.y
	98	84.8	4.5	500	10	AW490389 AW490389 UI-M-BH3
	99	84.8	4.5	574	9	AI171721 AI171721 EST217703
	100	84.8	4.5	588	10	AW491078 AW491078 UI-M-BH3
	101	84.8	4.5	655	10	AV647680 AV647680
	102	84.8	4.5	963	14	BQ936063 BQ936063 AGENCOURT
	103	84.2	4.4	466	9	AA744644 AA744644 ny79102.s
c	104	84.2	4.4	610	14	BM941929 BM941929 UI-M-CG0p
	105	84.2	4.4	621	12	BF143483 BF143483 601790842
c	106	84.2	4.4	972	17	CNS04TUQ AL402471 Tetraodon
c	107	84	4.4	276	13	BI300572 BI300572 UI-R-CV2
c	108	84	4.4	955	17	CNS04QE6 AL402487 Tetraodon
	109	83.8	4.4	783	9	AL561358 AL561358 AL561358
	110	83.8	4.4	1103	14	BM908460 BM908460 AGENCOURT
	111	83.8	4.4	1114	12	BG031969 BG031969 602300663
c	112	83.6	4.4	714	17	CNS01VNY AL169351 Tetraodon
c	113	83.2	4.4	259	13	BI299473 BI299473 UI-R-CV2-
	114	83.2	4.4	750	13	BI759107 BI759107 603042512
	115	83.2	4.4	818	13	BM020032 BM020032 603648565
	116	82.8	4.4	534	9	AA517920 AA517920 vb95e10.r
	117	82.6	4.4	663	13	BI832037 BI832037 603075640
	118	82.6	4.4	682	13	BI827058 BI827058 603075729
	119	82.2	4.3	317	10	BB276737 BB276737
	120	82.2	4.3	431	10	BE665969 BE665969 RPCI23 MA

ALIGNMENTS

```
RESULT 1
BG722442
LOCUS       BG722442                 727 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602693671F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4825937 5',
            mRNA sequence.
ACCESSION   BG722442
VERSION     BG722442.1  GI:14091629
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10739  row: n  column: 18
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1..727
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4825937"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (qtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT       182 a    194 c    175 g    176 t
ORIGIN

  Query Match          25.6%;  Score 486.4;  DB 12;  Length 727;
  Best Local Similarity   99.4%;  Pred. No. 4.8e-128;
  Matches  589;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTCTGCAAGGTGTTTTCTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTCTGCAAGGTGTTTTCTCTA 274

Qy         61 GCTGTTATCTCAAGACATGGATATAAAA-AACTCACCATCTAGCCTTAATTCTCCTTCCT 119
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        275 GCTGTTATCTCAAGACATGGATATAAAAGAACTCACCATCTAGCCTTAATTCTCCTTCCT 334

Qy        120 CCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCT 179
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 CCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCT 394

Qy        180 CCTATGTACACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGA 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 CCTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGA 454

Qy        240 ATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTTCGGCAGACCACAA 299
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 ATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTTCGGCAGACCACAA 514

Qy        300 GCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGT 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 GCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGT 574

Qy        360 TATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAAC 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 TATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAAC 634

Qy        420 ACACCTTACCTGTAAACAGAGAGACACTGAAAA-GGAAGGTTAGTGGGAACCGTTGCGCC 478
              |||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||
Db        635 ACACCTTACCTGTAAACAGAGAGACACTGAAAACGGACGGTTAGTGGGAACCGTTGCGCC 694

Qy        479 AGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 510
              ||||||||||||||||||||||||||||||||
Db        695 AGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 726


RESULT 2
BM942391/c
LOCUS       BM942391                 627 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  UI-M-BH3-bsc-b-09-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-bsc-b-09-0-UI 5', mRNA sequence.
ACCESSION   BM942391
VERSION     BM942391.1  GI:19402254
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N 7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE
FEATURES             Location/Qualifiers
     source          1..627
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-bse-b-09-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage "27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI: The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately  derived from a mixture of individually tagged
                     normalized libraries from  ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland,  striatum,
                     hippocampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows:  PCR amplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as  a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of  single-stranded circles.  The
                     remaining single-stranded circles  (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B  bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S4 library.  This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
BASE COUNT      147 a    151 c    184 g    144 t      1 others
ORIGIN

  Query Match           11.0%;  Score 208;  DB 14;  Length 627;
  Best Local Similarity   63.3%;  Pred. No. 3.3e-48;
  Matches  335;  Conservative    0;  Mismatches  191;  Indels    3;  Gaps    1;

Qy       1052 GACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTCC 1111
              || || || |  ||  | ||| |   ||| ||||| |  | ||||| || ||  | |||
Db        624 GATCAGGTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGG 565

Qy       1112 CGCTCAATTGACCACCCGGCAAGCTCATCTTTGTCCAGATCTTGTTCTGGACAGGAC 1171
              ||||| || || ||||| |  |||||| | | ||||||  |  |  | |||||||||| ||
Db        564 CGCTCCATGGAACACCCGGACAAGCTCCTGTNTGCTCCTAACTTGCTCCTGGACAGGAAT 505

Qy       1172 GAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATCCTCCTGCCAACTACTTCA 1231
              | || ||||| || ||||| ||  |||| ||||||||||||| | ||||| ||  |
Db        504 CAAGGTAAATGTGTGGAAGGCATGGTGGACATCTTTGACATGTTGCTGGCTACGTCAAGT 445

Qy       1232 AGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1291
              |||| ||   | | || || ||     |||| | | | ||  |||| |||| ||  ||
Db        444 CGGTTCCGCATGATGAACCTGCAGGGAGAAGAGTTTGTGTGCCTCAAATCCATCATTTTG 385

Qy       1292 CTCAATTCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGG 1348
              || |||||| |  ||||| |   |||| ||   |      ||  |   ||    |
Db        384 CTTAATTCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGAC 325

Qy       1349 AAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGC 1408
              |  |    |  | ||| ||    | || ||  ||||| |     |||| |||||
Db        324 CACATCCACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCT 265

Qy       1409 GGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCAC 1468
              ||| |  |    ||||||||        ||||| ||| | ||||  |  | || |||||
Db        264 GGCCTGACCCCCCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTGTTGCCAT 205

Qy       1469 CTCAGGCATGCGACTAACAAGGCCATGGAACATCTGCTCAACATCAAGTGCAAAAATGTG 1528
              || ||||   |||||||||| |||||||| |||||   |||||||||| |||| || ||
Db        204 ATCCGGCACATGAGTAACAAAGCCATGGAGCATCTCTACAACATGAAATGCAAGAACGTT 145

Qy       1529 GTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1577
              || ||  | |||||||||||| |||||||| || ||||||||   ||||
Db        144 GTGCCCCTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 96


RESULT 3
BB666338
LOCUS       BB666338                652 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB666338 RIKEN full-length enriched, 2 days pregnant adult female
            ovary Mus musculus cDNA clone E330013E21 5', mRNA sequence.
ACCESSION   BB666338
VERSION     BB666338.1  GI:16499971
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sato,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap trapper-selected cDNAs to
            prepare full length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            Waga,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
```

```
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone "E330013E21"
                     /clone_lib="RIKEN full-length enriched, 2 days pregnant
                     adult female ovary"
                     /sex "female"
                     /tissue_type "ovary"
                     /dev_stage "2 days pregnant adult"
                     /lab_host "DH10B"
                     /note "Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       149 a    193 c    165 g    155 t
ORIGIN

  Query Match           9.9%;  Score 188.2;  DB 10;  Length 652;
  Best Local Similarity   79.5%;  Pred. No. 1.6e-42;
  Matches  236;  Conservative    0;  Mismatches   58;  Indels    3;  Gaps    1;

Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGTTGTTATCT---CAAGACATGGATAT 84
              | | ||||||||||||  |||||  ||||||||||||  | ||   ||||||||||| ||
Db        354 CATCTGTGCCTCTTCTTCACAAGGATTTTTCTCAGTTGAGACCTACGCAAGACATGGAGAT 413

Qy         85 AAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTT 144
               |||||||||||||| |||||||  ||| ||| | || || ||||| || || |||||| |
Db        414 CAAAAACTCACCATCAAGCCTTACTTCCCCTGCTTCTTATAACTGTAGCCAGTCCATCCT 473

Qy        145 ACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATA 204
              |||| |||||| ||  |||| || || ||||||||||||||||| ||||  || |||||
Db        474 ACCCTTGGAGCATGGTCCCATCTATATCCCTTCCTCCTATGTAGAGAGCCGTCACGAATA 533

Qy        205 TCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCAC 264
                |||||||||||||||| || |||||||||||||| |||||  |||||||||     |
Db        534 CTCAGCCATGACATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGC 593

Qy        265 TAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAAC 321
              | ||   | | |||||||||| ||| |||||  ||||||||||||| | ||||||||
Db        594 TTACCCTGGAAGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTGCTATGGCCAAC 650


RESULT 4
BB609187
LOCUS       BB609187                 666 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB609187 RIKEN full-length enriched, 2 days pregnant adult female
            ovary Mus musculus cDNA clone E330010N18 5', mRNA sequence.
ACCESSION   BB609187
VERSION     BB609187.2  GI:16451164
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanaqaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Soqabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Dec 6, 2000 this sequence version replaced gi:11564364.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanaqawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: qenome res@qsc.riken.qo.jp,
            URL:http://qenome.qsc.riken.qo.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Suqahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            qenes. Genome Res. . 10 (10), 1617-1640 (2000)
             waqi,K., Fujiwake,S., Inoue,K., Toqawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Suqahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.qsc.riken.qo.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Mus musculus"
                     /db_xref "taxon:10090"
                     /clone="E330010N18"
                     /clone_lib "RIKEN full-length enriched, 2 days pregnant
                     adult female ovary"
                     /sex="female"
                     /tissue_type="ovary"
                     /dev_stage="2 days pregnant adult"
                     /lab_host="DH10B"
                     /note "Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
```

GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC 1."
BASE COUNT 141 a 199 c 165 g 161 t
ORIGIN

Query Match 9.3%; Score 177.2; DB 10; Length 666;
Best Local Similarity 76.5%; Pred. No. 2.3e-39;
Matches 231; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

```
Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCT---CAAGACATGGATAT 84
              | | ||||||||||||  |||||  ||||||||||||  | ||   ||||||||||| ||
Db        351 CATCTGTGCCTCTTCTCACAAGGATTTTTCTCAGCTGAGACCTACGCAAGACATGGAGAT 410

Qy         85 AAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTT 144
               |||||||||||||| ||||||| ||| ||| | || || ||||| || || |||||| |
Db        411 CAAAAACTCACCATCAAGCCTTACTTCCCCTGCTTCTTATAACTGTAGCCAGTCCATCCT 470

Qy        145 ACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATA 204
              |||| ||||||| ||  |||| || || ||||||||||||||||| ||||  || |||||
Db        471 ACCCTTGGAGCATGGTCCCATCTATATCCCTTCCTCCTATGTAGAGAGCCGTCACGAATA 530

Qy        205 TCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCAC 264
              ||||||||||||||||| || |||||||||||||| |||||  |||||||||     |
Db        531 CTCAGCCATGACATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGC 590

Qy        265 TAACTTGGAAGGTGGGCCTGGTCGGCAGACCAACAGCCCAAATGTGTTGTCCCCAACAGC 324
              |||| | |||||  || | ||  ||    | ||||||||||||| | | ||||    |
Db        591 TAACCTTGAAGGGTGGGCCTGTTCGCCAGACTGAAGCCCAAATGTGCTATTGCCACTTTC 650

Qy        325 TC 326
              ||
Db        651 TG 652
```

RESULT 5
BB666608
LOCUS BB666608 627 bp mRNA linear EST 26-OCT-2001
DEFINITION BB666608 RIKEN full-length enriched, 2 days pregnant adult female ovary Mus musculus cDNA clone E330023F08 5', mRNA sequence.
ACCESSION BB666608
VERSION BB666608.1 GI:16500241
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 627)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kondo,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki Y., Okido,T., Saito,R., Sakai,C., Sakai,E., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro cho, Tsurumi ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap trapper-selected cDNAs to prepare full length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1..627
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone "E330023F08"
/clone_lib="RIKEN full-length enriched, 2 days pregnant adult female ovary"
/sex="female"
/tissue_type="ovary"
/dev_stage="2 days pregnant adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTIVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1."
BASE COUNT 133 a 185 c 161 g 148 t
ORIGIN

Query Match 9.3%; Score 176.4; DB 10; Length 627;
Best Local Similarity 80.3%; Pred. No. 3.8e-39;
Matches 220; Conservative 0; Mismatches 51; Indels 3; Gaps 1;

```
Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCT---CAAGACATGGATAT 84
              | | ||||||||||||  |||||  ||||||||||||  | ||   ||||||||||| ||
Db        354 CATCTGTGCCTCTTCTCACAAGGATTTTTCTCAGCTGAGACCTACGCAAGACATGGAGAT 413

Qy         85 AAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTT 144
               |||||||||||||| ||||||| ||| ||| | || || ||||| || || |||||| |
Db        414 CAAAAACTCACCATCAAGCCTTACTTCCCCTGCTTCTTATAACTGTAGCCAGTCCATCCT 473

Qy        145 ACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATA 204
              |||| ||||||| ||  |||| || || ||||||||||||||||| ||||  || |||||
Db        474 ACCCTTGGAGCATGGTCCCATCTATATCCCTTCCTCCTATGTAGAGAGCCGTCACGAATA 533

Qy        205 TCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCAC 264
              ||||||||||||||||| || |||||||||||||| |||||  |||||||||     |
Db        534 CTCAGCCATGACATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGC 593

Qy        265 TAACTTGGAAGGTGGGCCTGGTCGGCAGACCACA 298
              |||| ||||||||||||||| ||| |||||  ||
Db        594 TAACCTTGAAGGTGGGCCTGTTCGCCAGACTGCA 627
```

RESULT 6

```
BM933826/c
LOCUS       BM933826                 528 bp    mRNA    linear   EST 13-MAR 2002
DEFINITION  UI_M_BH3_bsc_b_09_0_UI.r2 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI_M_BH3_bsc_b_09_0_UI 5', mRNA sequence.
ACCESSION   BM933826
VERSION     BM933826.1  GI:19392978
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism "Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref "taxon:10090"
                     /clone="UI_M_BH3_bsc_b_09_C_UI"
                     /clone_lib "NIH_BMAP_M_S4"
                     /dev_stage "27-32 days"
                     /lab_host "DH10B (Life Technologies)"
                     /note "Vector: pT7T4D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately  derived from a mixture of individually tagged
                     normalized libraries from  ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland,  striatum,
                     hippocampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows:  PCR-amplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as  a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of  single stranded circles.  The
                     remaining single-stranded circles  (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double stranded circles and electroporated
                     into DH10B  bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S4 library.  This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
BASE COUNT      119 a    123 c    156 g    128 t      2 others
ORIGIN

  Query Match             8.6%;  Score 162.4;  DB 14;  Length 528;
  Best Local Similarity   62.7%;  Pred. No. 4.6e-45;
  Matches  269;  Conservative    0;  Mismatches  157;  Indels    3;  Gaps    1.

Qy       1152 ATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGAAATCTTTGACA 1211
             |  |  | ||||||||| || | || ||||| || ||||| ||  |||| ||||||||||
Db        525 ACNTGCTCCTGGACAGGAATCAAGGTAAATGTGTGGAAGGCATGGTGGAGATCTTGACA 466

Qy       1212 TGCTCCTCCCAACTACTTCAACGTTTCGACAGTTAAAACTCCAACACAAACAATATCTCT 1271
             || | |||| || |    |||| ||   | | || || ||     |||| | | | |
Db        465 TGTTGCTGGCTACGTCAAGTCGGTTCCGCATGATGAACCTGCAGGGAGAAGAGTTTGTGT 406

Qy       1272 GTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCC-  TCTGGTCACAGCGACCC 1328
             |  |||| |||| ||  |||| |||||| |  ||||| |   |||| ||   |
Db        405 GCCTCAAATCCATCATTTTGCTTAATTCCGGAGTGTACACGTTTCTGTCCAGCACCTTGA 346

Qy       1329 AGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGG 1388
             || |  ||   |    |  |   |  ||| ||    | || ||  |||||
Db        345 AGTCTCTGGAAGAGAAGGACCACATCCACCGTGTCCTGGACAAGATCACAGACACTTTGA 286

Qy       1389 TTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACC 1448
             |    |||| |||||   ||| |  |   ||||||||      ||||| ||| | |
Db        285 TCCACCTGATGGCCAAAGCTGGCCTGACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGC 226

Qy       1449 TCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCA 1508
             ||||  |  | || |||||  || ||||  |||||||| |||||||| |||||  ||
Db        225 TCCTTCTCATTCTTTCCCATATCCGGCACATGAGTAACAAAGGCATGGAGCATCTCTACA 166

Qy       1509 ACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCC 1568
             ||||||| |||| || || || || | |||||||||| |||||||| || ||||||
Db        165 ACATGAAATGCAAGAACGTTGTGCCCCTCTATGACCTGCTCCTGGAGATGTTGGATGCCC 106

Qy       1569 ACGTGCTTC 1577
             ||   ||||
Db        105 ACCGCCTTC 97


RESULT 7
BG738138
LOCUS       BG738138                 580 bp    mRNA    linear   EST 15-MAY 2001
DEFINITION  tp04105.y1 zebrafish qridded kidney Danio rerio cDNA clone 4728705
            5' similar to TR:Q13012 Q13012 ESTROGEN RECEPTOR. ;, mRNA sequence.
ACCESSION   BG738138
VERSION     BG738138.1  GI:14087827
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvoy,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     other_ESTs: tp04105.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            Genome Systems, St. Louis, Missouri (web address:
            www.genomesystems.com) (email contact: info@genomesystems.com) and
            Research Genetics, Huntsville, Alabama (web address: www.resgen.com
            ) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .580
```

```
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone "4728705"
                     /clone_lib="zebrafish gridded kidney"
                     /sex="mixed"
                     /tissue_type "kidney pooled from 300 wild type adults"
                     /lab_host="XLOLR"
                     /note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI;
                     Site_2: XhoI; oligo dT cDNA library constructed from mRNA
                     pooled from pooled kidney  tissue from 300 adult
                     zebrafish."
BASE COUNT      133 a    161 c    156 g    130 t
ORIGIN

  Query Match             8.5%;  Score 162.2;  DB 12;  Length 580;
  Best Local Similarity   76.0%;  Pred. No. 4.3e-35;
  Matches  200;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         503 AAGAGGGAAGCTCACTTCTGGGCTGTCTGCAGCGATTACGCATCGGCATATCACTATCCCA 562
               ||  | ||  |||||||||||| || |||   || ||||| || || ||||||||||||
Db         257 AAAGGAGACACTCACTTCTGTGCGCGGTGTGCCATGACTACGCTTCTGGGTATCACTATCGT 316

Qy         563 GTCTGGTCGTGTGAAGGATGTAAGCCCTTTTTTAAAAGAAGCATTCAAGGACATAATCAT 622
               |||||||| |||||||| || ||||| || || || || | |||||||||||| || |||||
Db         317 GTCTGGTCATGTGAAGGGTGCAAGCCTTTCTTCAAGCGTAGCATTCAAGGTCACAATCAC 376

Qy         623 TATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAG 682
               |||||||||||||| || || ||||| || || || || | | | |||||||||||| |||
Db         377 TATATTTGTCCAGCCACCAACCAGTGCACTATTCACAAGAGCAGACGCAAGAGCTGTCAG 436

Qy         683 GCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGCGAAGTGTGGCTCCCGCACAGAC 742
               ||||| |||| || |||| || |||||||| ||| ||||||||||     |||| |||
Db         437 GCCTGTCGACTCCGCAAGTGCTATGAAGTGGGCATGATGAAGTGTGGTGTTAGGAGGGAG 496

Qy         743 AGATGTGGGTACCGCCTTGTGCG 765
                 | ||  | || ||   ||  ||
Db         497 CGCTGCAGTTATCGAGGTGCTCG 519


RESULT 8
AA224131/c
LOCUS       AA224131                 531 bp    mRNA    linear   EST 11 MAR 1998
DEFINITION  zr14e02.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
            IMAGE:648794 5', mRNA sequence.
ACCESSION   AA224131
VERSION     AA224131.1  GI:1844690
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1309  Std Error: 0.00
            Seq primer: -28ml3 revl ET from Amersham
            High quality sequence stop: 357.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Homo sapiens"
                     /db_xref="GDB:5589314"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:648794"
                     /clone_lib="Stratagene hNT neuron (#937233)"
                     /dev_stage="hNT neurons"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note "Vector: pBluescript SK ; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                     Differentiated, post mitotic hNT neurons. Average insert
                     size: 1.5 Kb. Gni ZAP XR Vector;  5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3'  3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      192 a     74 c    112 g    146 t      7 others
ORIGIN

  Query Match             8.3%;  Score 158;  DB 9;  Length 531;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-34;
  Matches  158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         184 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 125

Qy          61 GCTCTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         124 GCTCTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 65

Qy         121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACG 158
               ||||||||||||||||||||||||||||||||||||||
Db          64 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACG 27


RESULT 9
BF058527
LOCUS       BF058527                 642 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  7k33h08.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3477450 3'
            similar to SW:ESR1_HUMAN P03372 ESTROGEN RECEPTOR ;, mRNA sequence.
ACCESSION   BF058527
VERSION     BF058527.1  GI:10812423
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Hento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 492.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3477350"
                     /clone_lib="NCI_CGAP_Ov18"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCGACATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
```

```
                    (Pharmacia), digested with Not I and cloned into the Not
                    I and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       158 a    182 c    196 g    106 t
ORIGIN

  Query Match             8.3%;  Score 157;  DB 12;  Length 642;
  Best Local Similarity   77.6%;  Pred. No. 1.4e-33;
  Matches  190;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        346 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 405

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATCATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        406 TGCTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGACAAGTATTCAAGGACATAACGACTAT 465

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || ||||| || || |||||||||| || | ||||||||||||||||
Db        466 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 525

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db        526 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 585

Qy        746 TGTGG 750
                | ||
Db        586 AGAGG 590


RESULT 10
AW144052
LOCUS       AW144052                 782 bp    mRNA    linear   EST 27-OCT-1999
DEFINITION  fi14h07.y1 Suqano Kawakami zebrafish DRA Danio rerio cDNA clone
            2601181 5' similar to qb:M12674 ESTROGEN RECEPTOR (HUMAN);, mRNA
            sequence.
ACCESSION   AW144052
VERSION     AW144052.1  GI:6145659
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopteryqii; Neopteryqii; Teleostei; ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 782)
  AUTHORS   Suqano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,P.
  TITLE     Wash0 Zebrafish EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     other_ESTs: fi14h07.x1
            Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Suqano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1..782
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="2601181"
                     /clone_lib "Suqano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eqqs)"
                     /dev_staqe="adult"
                     /lab_host "DH10B (phaqe resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oliqo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was liqated to a DraIII adaptor
                     [TGTTGGCCTACTGG], diqested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 4' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fraqments
                     <1.5kb. Library constructed by Dr. Sumio Suqano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT       195 a    230 c    176 g    179 t      2 others
ORIGIN

  Query Match             8.3%;  Score 157;  DB 10;  Length 782;
  Best Local Similarity   59.4%;  Pred. No. 1.6e-33;
  Matches  325;  Conservative    0;  Mismatches  207;  Indels   15;  Gaps    3;

Qy        144 TACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAAT 203
              | ||  |||| | |  | |||| | ||| || ||  ||||  ||||    ||| || |
Db        251 TGCCAGTGGAGAATCACCCCATCTGCATCCCATCGCCCTACACAGACCTTGGCCACGACT 310

Qy        204 ATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCA 263
                  | || | || || || || || ||| || ||   |||||||   |
Db        311 TCAGCACTCTGCCCTTTTACAGTCCCGCTCTGCTGGGGTACAGCACATCGC   -- --- 361

Qy        264 CTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACAC 323
              ||  | ||| |  | |||  || ||| |  |||||| | | | || |||||| | |
Db        362 CTTTATCGGACTGCTCGTCTGTGCGCCAGTCGCTAAGCCCGACTTTATTTTGGCCACCTC 421

Qy        324 CTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTC 383
                | ||  |||| |  |    | || |  |||    || ||   |   | ||
Db        422 ACAGCCATGTTTCCTCACTCACATTGCAACAACAGAGTCGACTTCAACAGAACCATGCTA 481

Qy        384 AAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGA 443
                |   | ||||   |||  | | | |  || ||      |   | ||||||  |
Db        482 CTAGTGGGACTTGGACAGAACATACACCACATGATCATGTTGAAGAGGAAAACAGCAAAC 541

Qy        444 CACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAA 503
              |||||  | |  || ||| |  || | | | |  |  |||  |||      |   |
Db        542 CACTGGTGAAG-CGGGTAGCAGATACAGAGGAGACGTCTGTGTCTCTGA-----GAGGTA 595

Qy        504 AGAGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAG 563
              |    ||    |||| ||| || |||||||| |||||||| || |||||||| |||||||
Db        596 AAGCTGACATGCACTACTGTGCCGTCTGCAGTGATTACGCCTCTGGATATCATIATGGAG 655

Qy        564 TCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATCATT 623
              | |||||||||||||| || || |||||||| || || || || |||||||| ||||| |
Db        656 TGTGGTCGTGTGAAGGCTGCAAAGCCTTTTTCAAGAGGAGTATACAAGGACACAATGACT 715

Qy        624 ATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGG 683
              | || ||||| || ||||| || || || |||||  | ||||||||||  ||||||||||
Db        716 ACATCTGTCCTGCCACAAACCAATGCACTATCGANCAGAACCGGCGCANAAGCTGCCAGG 775

Qy        684 CCTGCCG 690
              |||||||
Db        776 CCTGCCG 782


RESULT 11
AW976636
LOCUS       AW976636                 641 bp    mRNA    linear   EST 02-JUN-2000
DEFINITION  EST388745 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW976636
VERSION     AW976636.1  GI:8167867
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 363
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGN"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      159 a    175 c    195 g    112 t
ORIGIN

  Query Match             8.2%;  Score 155.4;  DB 10;  Length 641;
  Best Local Similarity   77.1%;  Pred. No. 4.1e-33;
  Matches  189;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATCGACTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        334 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATCGAGTC 393

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        394 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACCACTAT 453

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || ||||| || || ||||||||| || | ||||||||||||||||
Db        454 ATGTGTCCAGCCACCAACCAGTGCCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGCCC 513

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db        514 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGCCATACCAAAACACCGA 573

Qy        746 TGTGG 750
              |||
Db        574 ATAGG 578


RESULT 12
BF660250
LOCUS       BF660250                 648 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  maa95q12.x1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3824830 3' similar to SW:ESR1_MOUSE P19785 ESTROGEN RECEPTOR
            ;, mRNA sequence
ACCESSION   BF660250
VERSION     BF660250.1  GI:11925456
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1461182
            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3824830"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type "Thymus"
                     /dev_stage "4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      158 a    184 c    184 g    123 t
ORIGIN

  Query Match             8.1%;  Score 153.8;  DB 12;  Length 648;
  Best Local Similarity   76.7%;  Pred. No. 1.2e-32;
  Matches  188;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATCGACTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        283 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATCGAGTC 342

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| || ||||| || ||||| || ||||| ||||||||||||||||| ||||| ||
Db        343 TGGTCCTGTGCAAGGGTGCAAGGCTTTCTTTAAGAGAAGCATTCAAGGACACAATGACTAC 402

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||||||||| || || || || || || |||||| | ||||| |||||||||
Db        403 ATGTGTCCAGCTACAAACCAATGCCACCATTGACAAGAACCGCAGGAAGAGCTGCCAGCCC 462

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              || || || || ||||||||||||||||| ||| |||| | |||  |||| ||| |
Db        463 TGTCGGCTGGGCAAGTGTTACGAACTGGGCATGATGAAAGGTGCCATACCAAAGACCGC 522

Qy        746 TGTGG 750
              | ||
Db        523 CGAGG 527


RESULT 13
BF733032
LOCUS       BF733032                 642 bp    mRNA    linear   EST 30-MAR-2001
DEFINITION  nae18e05.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3435729 3'
            similar to SW:ESR1_HUMAN P03372 ESTROGEN RECEPTOR ;, mRNA sequence.
ACCESSION   BF733032
VERSION     BF733032.1  GI:12058107
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
```

```
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:3435729"
                     /clone_lib "NCI_CGAP_Ov18"
                     /tissue_type "fibrothecoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note "Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCGACATTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Benito Soares and M. Fatima Bonaldo. "
BASE COUNT      156 a    184 c    193 g    109 t
ORIGIN

  Query Match             7.9%;  Score 149.6;  DB 12;  Length 642;
  Best Local Similarity   77.1%;  Pred. No. 1.9e-31;
  Matches  182;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        446 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 405

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        406 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 465

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || ||||| || || |||||||||| || | ||||||||||| |||
Db        466 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAAGCC 525

Qy        686 TGCCGACTTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGA 741
              ||||| || || || || |||||| |||||||| |||| |||| ||   |||
Db        526 TGCCGGCTCCGCAAATGCTACGAAATGGGAATGATGAAAGGTGGGATACGATAAGA 581


RESULT 14
BG084311
LOCUS       BG084311                 594 bp    mRNA    linear   EST 26-JAN-2001
DEFINITION  H3099A08 5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3099A08 5', mRNA sequence.
ACCESSION   BG084311
VERSION     BG084311.1  GI:12566875
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Karqul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3099A08-3
            Contact: George J. Karqul
            Laboratory of Genetics
            National Institute on Aqinq/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224 6820, USA
            Email: cdna@lqsun.qrc.nia.nih.qov
            This clone set has been freely distributed to the community. Please
            visit http://lqsun.qrc.nia.nih.qov/cDNA/15k.html for details.
            Plate: H3099  row: A  column: 08
            Seq primer: -21M13 Reverse
            High quality sequence stop: 594
            POLYA-No.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3099A08-5"
                     /db_xref="taxon:10090"
                     /clone "H3099A08"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage "Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation staqe
                     embryos from  unfertilized eqq to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/qonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with oliqo(dT
                     ) Not primers.  References include: (1) Genome-wide
                     expression profiling of mid qestation  placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray,  2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127 9132; (2)
                     Large-scale cDNA  analysis reveals phased qene expression
                     patterns during preimplantation  mouse develolpment, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide  mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day  mouse embryos reveals enrichment in the t-complex
                     and under-representation  on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT      170 a    124 c    167 g    133 t
ORIGIN

  Query Match             6.9%;  Score 131.4;  DB 12;  Length 594;
  Best Local Similarity   79.2%;  Pred. No. 3e-26;
  Matches  156;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy        554 CACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGA 613
              || ||||| |||||||| || ||||| || ||||| || ||||| ||||||||||||||||
Db          1 CATTATGGGGTCTGGTCCTGCGAAGGCTGCAAGGCTTTCTTTAAGAGAAGCATTCAAGGA 60

Qy        614 CATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAG 673
              || ||||| || || |||||||||||||| || || || || || || || |||||| | |||
Db         61 CACAATGACTACATGTGTCCAGCTACAAACCAATGCACCATTGACAAGAACCGGAGGAAG 120

Qy        674 AGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCC 733
              || |||||||||| || || || |||||||||||||||| ||| |||| | |||
Db        121 AGTTGCCAGGCCTGTCGGCTGCGCAAGTGTTACGAAGTGGGCATGATGAAAGGCGGCATA 180

Qy        734 CGGAGAGAGAGATGTGG 750
              |||| ||| | | | ||
Db        181 CGGAAAGACCGCCGAGG 197


RESULT 15
BM810012
LOCUS       BM810012                 1142 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6581396 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454957
            5', mRNA sequence.
ACCESSION   BM810012
VERSION     BM810012.1  GI:19126835
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1142)
```

```
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cqapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1950  row: o  column: 22
            High quality sequence start: 13
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .1142
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454957"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      191 a    334 c    398 g    215 t      4 others
ORIGIN

  Query Match             6.8%;  Score 129.8;  DB 14;  Length 1142;
  Best Local Similarity   48.7%;  Pred. No. 1.2e-25;
  Matches  381;  Conservative    0;  Mismatches  399;  Indels    3;  Gaps    1;

Qy        517 CTTCTGGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGA 576
              | |||||   |||||  | ||    || || || || |||||||| ||    || |||||
Db         44 CCTCTGCCTGGTCTGTGGGGACGTGGGCTCCGGCTACCACTATGGTGTGGCATCCTGTGA 103

Qy        577 AGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGC 636
              | || || ||||| || || || | ||| || ||    |  || || |  |||| ||
Db        104 GGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGCAGCATCGAGTACAGCTGTGCCCC 163

Qy        637 TACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCC 696
              | ||  |||||   |||   ||     | ||||||  ||||||||||||||||  |
Db        164 CTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCAC 223

Qy        697 CAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCG 756
              |||||       ||||| ||| | |||   ||     ||   ||  |     ||    |
Db        224 CAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCCCCTGC 283

Qy        757 CCTTGTGCGGAGACAGAGAAGTGCCGACGAGCACCTGCACTGTGCCGGCAAGGCCAACAG 816
              |   |  |   || |  | | || | | | |  |     |        ||
Db        284 GCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCCC 343

Qy        817 AAGTGCGGGCCAGGCCCCCCCAGTGCGGGAGCTGCTGCTGGAGGCCTGACCCCCAACCA 876
              |||   ||  ||   |  |  |    | | |  |  ||||        | |
Db        344 TGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACT 403

Qy        877 GCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGCCCCAGTGC 936
              | |    |   ||  ||| |      |  |        ||      ||||   |
Db        404 GGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGC 463

Qy        937 GCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTCGT 996
              ||        |   | |     ||    ||||     | |    |||  ||| | ||
Db        464 CCCTGATGGGCACCTCCAGCCGTGGCTACCTCTGTGACCTCTTTGACCGAGAGATTGT 523

Qy        997 ACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCA 1056
              ||  ||||||||||||||||  || || |||||      |||    ||||  |||||
Db        524 GGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCA 583

Qy       1057 AGTGCGGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTC 1116
                ||    |  || ||||   ||||||||||| |  || ||||  ||     ||||||
Db        584 GATGTCAGTACTGCACAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGCTC 643

Qy       1117 AATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGG 1176
              | |    |  | |   ||||   ||| |||   ||  | || ||||    || |||||
Db        644 ACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTAGTCCTGGA---TGAAGAGGG 700

Qy       1177 GAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTT 1236
              | | | | ||  ||  |   |||| |    |      || ||| |          || |
Db        701 GGCACGGGCAGCTGGCCTGNGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGCGGCT 760

Qy       1237 TCGACAGTTAAAACTCCAACACAAACAATATCTCTGTGTCAAGGCCATGATGCTGCTCAA 1296
              | |   |    || | |   |||| ||| |      |  |||| ||  |||   |||
Db        761 GCAGGCCCTGCGGCTGGAGCGAGAAGAGTATGTCTACTANAGGCCTTGGCCCTTGCCAA 820

Qy       1297 TTC 1299
              |||
Db        821 TTC 823


RESULT 16
AF455798
LOCUS       AF455798                 137 bp    DNA     linear   GSS 23-APR-2002
DEFINITION  AF455798 Macaca mulatta DNA (Murphy WJ) Macaca mulatta genomic
            clone ESR2, DNA sequence.
ACCESSION   AF455798
VERSION     AF455798.1  GI:20271396
KEYWORDS    GSS.
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 137)
  AUTHORS   Murphy,W.J., Page,J.E., Smith,C. Jr., Desrosiers,R.C. and O'Brien
            ,S.J.
  TITLE     A radiation hybrid mapping panel for the rhesus macaque
  JOURNAL   J. Hered. 92 (6), 516-519 (2001)
  MEDLINE   21945527
COMMENT     Contact: Murphy WJ
            Laboratory of Genomic Diversity
            National Cancer Institute-FCRDC
            P.O. Box B, Bldg. 560, Rm.21-10, Frederick, MD 21702, USA
            Tel: 301 846 1299
            Fax: 301 846 6327
            Email: murphywi@mail.ncifcrf.gov
            Class: unknown.
FEATURES             Location/Qualifiers
     source          1. .137
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /clone="ESR2"
                     /clone_lib="Macaca mulatta DNA (Murphy WJ)"
BASE COUNT       34 a     40 c     30 g     33 t
ORIGIN

  Query Match             6.7%;  Score 126.4;  DB 17;  Length 137;
  Best Local Similarity   95.6%;  Pred. No. 3.7e-25;
  Matches  130;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        240 ATTACAGCAGTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACGCAA 299
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          1 ATTACAGCAGTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACGGCAA 60

Qy        300 GCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGT 359
              |||||||||||||||||||||||||||| ||||||||||||||  ||||||||||||
Db         61 GCCCAAATGTGTTGTGGCCAACACCTCGACACCTTTCTCCTTTAGCACTCCATCGCCAGT 120

Qy        360 TATCACATCTGTATGC 375
              |||||||||| |||||
Db        121 TATCACATCTATATGC 136
```

```
RESULT 17
BG891892/c
LOCUS       BG891892                 558 bp    mRNA    linear   EST 04-JUN-2001
DEFINITION  fp46e11.xl zebrafish gridded kidney Danio rerio cDNA clone 4759773
            3' similar to TR:Q9W669 Q9W669 ESTROGEN RECEPTOR BETA. ;, mRNA
            sequence.
ACCESSION   BG891892
VERSION     BG891892.1  GI:14286502
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Thelsing,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            Genome Systems, St. Louis, Missouri (web address:
            www.genomesystems.com) (email contact: info@genomesystems.com) and
            Research Genetics, Huntsville, Alabama (web address: www.resgen.com
            ) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1..558
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="4759773"
                     /clone_lib="zebrafish gridded kidney"
                     /sex="mixed"
                     /tissue_type="kidney pooled from 300 wild type adults"
                     /lab_host="XL0LR"
                     /note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI;
                     Site_2: XhoI; oligo dT cDNA library constructed from mRNA
                     pooled from pooled kidney  tissue from 300 adult
                     zebrafish."
BASE COUNT      137 a    119 c    156 g    146 t
ORIGIN

  Query Match             6.4%;  Score 121;  DB 12;  Length 558;
  Best Local Similarity   68.9%;  Pred. No. 2.8e-23;
  Matches  166;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       1462 TGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTCATTGCCAAGAGCGGCATCTCCTCCC 1421
             ||||| || | ||||| || ||| |||||||||  ||  | |||| |||| |   || ||
Db        554 TGCTGGACTCGGTGACTGACGCTCTGGTTTGGGCCATCTCGAAGACCGGCCTTAGCTTCC 495

Qy       1422 AGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGA 1481
             ||||||  ||||   | ||||| | || || ||||| || || ||| |  | || |  |
Db        494 AGCAGCGCTCCACCAGACTGGCCCATCTGCTCATCCTGCTCTCAACACATACGCCACGTCA 435

Qy       1482 GTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATG 1541
             ||||||| |||||||| || |||| |  ||||||   || ||  |||| ||  ||||||
Db        434 GTAACAAAGGCATGGACCACCTGCACTGCATGAAGATGAAGAACATGGTTCCTCTGTATG 375

Qy       1542 ACCTGCTGCTGGAGATGCTGAATGCCCACCGTGCTTCGCGGGTGCAAGTCCTCCATCACGG 1601

             ||||||| |||||||||||| | || ||| |  | | | | | |   |  ||   | | |
Db        474 ACCTGCTCCTGGAGATGCTGGACGCTCACATCATGCACAGCTCCCGTTTGTCTCACTCCG 415

Qy       1602 G 1602
              |
Db        414 G 414


RESULT 18
BJ068905
LOCUS       BJ068905                 692 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ068905 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL051b20 5', mRNA sequence.
ACCESSION   BJ068905
VERSION     BJ068905.1  GI:17497655
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..692
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL051b20"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
                     /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett  and A.M. Zorn,
                     (Wellcome/CRC Institute). "
BASE COUNT      160 a    168 c    182 g    182 t
ORIGIN

  Query Match             6.2%;  Score 117;  DB 14;  Length 692;
  Best Local Similarity   51.6%;  Pred. No. 4.4e-22;
  Matches  329;  Conservative    0;  Mismatches  290;  Indels   18;  Gaps    2;

Qy        540 ACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAA 599
              | ||||| || ||||||||||| ||    || ||||| | ||| || ||||| || ||
Db         10 AGGCATCAGGCTATCACTATGGGGTTGCCTCATGTGAGGCATGCAAAGCCTTCTTCAAGC 69

Qy        600 GAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATA 659
              | | ||| ||||  ||| || || | |||||||| |  ||| | ||| ||| |  |
Db         70 GCACCATCCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATGAATGTGAAATTACCA 129

Qy        660 AAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGG 719
              | |  || || |||  ||| || || ||||  |    |||||       ||||| |||
Db        130 AGAGACGTCGAAAGGCCTGTCAAGCTTGCCGCTTCACAAAGTGCCTTCGCGTGGGCATGC 189

Qy        720 TGAAGTGTGGCTCCCCGCAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTG 779
              | ||    ||    ||    ||  |     |    | | |  |       |||| ||
Db        190 TAAAAGAAGGGGTACGCTTGGACCGTGTCAGAGGAGGACGTCAGAAATACAAGAGGAGGC 249

Qy        780 CCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCGAG 839
              | || |  | | | |  |                |     ||| || || || | | ||
Db        250 CAGATGGAGATATCCTC----------CAGTACTCGTCTGGTGGGCAGGCTCAGCAAG 297
```

```
Qy        840 TGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGG 899
              |     || |  |   |  |  || | |   |||  || | | ||||  ||
Db        298 CCTCTGCAACAGTGGTCAAGAAGCAAACTCCAGTGAATGCAGTTGTCTCTCACCTGCTGG 357

Qy        900 CTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATG- 958
               ||     ||  |   | || |       ||     |  |  |  | || |       ||
Db        358 TTGCAGAACCAGACAAGTTGTTTGCTATGCCAGACCAGCACTTCCTGATGGTGATTTCA 417

Qy        959 -----ATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGG 1013
                  ||||  |||||    |  | || ||    ||  | ||  ||| || |||||||
Db        418 AATATATGAGCACCCTGTGTGACCTTGCTGATCGTGAAATTGTCATCATCATTAGCTGGC 477

Qy       1014 CCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGA 1073
              | ||||| || || || ||        ||   ||| |  ||||  || | |||||| |||
Db        478 CAAAGAATATCCCAGGATTCTCCTCCCTTTCCCTCTCGGACCAGATGAGTCTCTTGCAGA 537

Qy       1074 GCTGTTGGATGGAGGTGTTAATGATGCGGCTGATGTGCCGCTCAATTGACCACCCCGCCA 1133
              |    |||||||||||| ||| || | ||  |  | |  ||||  ||    |    |
Db        538 GTGTCTGGATGGAGCTTTTATTGCTTGGACTTGTTTCCGCTCGCTTCCGTATGAAGATG 597

Qy       1134 AGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGA 1170
              || |  | ||||||   ||| |||| || ||   |||
Db        598 AGGTTGTGTTTGCTGAGGATTTTGTGCTTGATGAGGA 634


RESULT 19
AZ596395
LOCUS       AZ596395                 532 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0409D10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0409D10 R, DNA sequence.
ACCESSION   AZ596395
VERSION     AZ596395.1  GI:11718585
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M , Meenen,E , Pederson,T , Reilly
            ,M., Rose,M , Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0409  row: D  column: 10
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 532.
FEATURES             Location/Qualifiers
     source          1..532
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0409D10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares.html). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      118 a    132 c    135 g    147 t
ORIGIN

  Query Match           6.1%;  Score 115.6;  DB 17;  Length 532;
  Best Local Similarity   84.4%;  Pred. No. 9.6e-22;
  Matches  130;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       1165 CAGCGATCAGCGGAAATGCGTAGAAGGAATTCTGCAAAICTTTGACATGCTCCTGGCAAC 1224
              |||||||||||||| |||| |||| |||||||||||||||||||||||||||||||| ||
Db        257 CAGGGATGAGGGGAAGTGCGTGGAAGGATTCTGGAAATCTTTGACATGCTCCIGGCGAC 316

Qy       1225 TACTTCAAGGTTTCCAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCAT 1284
              ||  || |||| || |||||||||||| || ||||||||||||| |||| |||||||
Db        317 GACGGCACGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCAT 376

Qy       1285 GATCCTGCTCAATTCCAGTATGTACCCTCTGGTC 1318
              ||| || |||| |||||  ||| || |  | |
Db        377 GATTCTCCTCAACTCCAGTGAGTAGCCACGAGGC 410


RESULT 20
AQ719425
LOCUS       AQ719425                 611 bp    DNA     linear   GSS 14-JUL-1999
DEFINITION  HS_5512_B2_D07_T7A RPCI 11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1088 Col-14 Row-H, DNA sequence.
ACCESSION   AQ719425
VERSION     AQ719425.1  GI:5479094
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderina_bac.htm)
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.hts.washington.edu
            Plate: 1088  row: H  column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 611.
FEATURES             Location/Qualifiers
     source          1..611
```

```
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "Plate:1088 Col 14 Row H"
                     /clone_lib "RPCI-11 Human Male BAC Library"
                     /sex "male"
                     /note "Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       178 a    141 c    129 g    144 t     19 others
ORIGIN

  Query Match              6.1%;  Score 115.2;  DB 17;  Length 611;
  Best Local Similarity   97.5%;  Pred. No. 1.3e-21;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        610 AGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCG 669
              | ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        337 AAGACATAATGATGATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCG 396

Qy        670 CAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 729
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        397 CAAGAGCTGCCAGGCCTGCCGACTTCTGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 456


RESULT 21
BC008218
LOCUS       BC008218                1443 bp    mRNA    linear   HTC 12-JUL-2001
DEFINITION  Homo sapiens, clone IMAGE:3871297, mRNA.
ACCESSION   BC008218
VERSION     BC008218.1  GI:14713363
KEYWORDS    HTC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1443)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 13 Row: k Column: 10
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4335517
            This clone has the following problem: incomplete processing.
FEATURES             Location/Qualifiers
     source          1..1443
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:3871297"
                     /tissue_type "Eye, retinoblastoma"
                     /clone_lib "NIH_MGC_67"
                     /lab_host "DH10B"
                     /note="Vector: pCMV SPORT6"
BASE COUNT       390 a    339 c    352 g    362 t
ORIGIN

  Query Match              5.8%;  Score 110.6;  DB 11;  Length 1443;
  Best Local Similarity   49.4%;  Pred. No. 4.5e-20;
  Matches  342;  Conservative    0;  Mismatches  324;  Indels   27;  Gaps    1;

Qy        527 GTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAG 586
              || ||  | ||   ||| || || || |||||||| ||    || |||||||  || |||
Db        732 GTGTGTGGTGACATCGCTTCTGGGTACCACTATGGGGTAGCATCATGTGAAGCCTGCAAG 791

Qy        587 GCCTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 646
              || || || || || | ||||||||  ||| || || | || || || 4| ||| |
Db        792 GCATTCTTCAAGAGGACAATTCAAGGCAATATAGAATACAGCTGCCCTGCCACGAATGAA 851

Qy        647 TGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTAC 706
              |||  |||| || | | || ||  |||||||||| |||| | ||||||||
Db        852 TGTGAAATCACAAAGCGCAGACGTAAATCCTGCCAGGCTTGCCGCTTCATGAAGTGTTTA 911

Qy        707 GAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGG 766
              ||||||| ||| |||| || || ||| |||
Db        912 AAAGTGGGCATGCTGAAAGAAGGGGTGCGTCTTGACAGA---------------- -- 950

Qy        767 AGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGC 826
                  | ||| |  |||| | ||  | | | |  |  |  ||
Db        951 ------GTACGTGGAGGTCGGCAGAAGTACAAGCGCAGGATAGATGCGGAGAACAGCCCA 1004

Qy        827 CACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTC 886
              ||  |  ||    || ||  | ||      |   |      | ||   |   |
Db       1005 TACCTGAACCCTCAGCTGGTTCAGCCAGCCAAAAAGCCATATAACAAGATTGTCTCACAT 1064

Qy        887 ACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACC 946
                | ||| |   | |              ||      ||| | ||  ||  ||
Db       1065 TTGTTGGTGGCTGAACCGGAGAAGATCTATGCCATGCCTGACCCTACTGTCCCCGACAGT 1124

Qy        947 GAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATC 1006
              ||   |       | |   | |||    | |||||||||  ||||||||    || ||
Db       1125 GACATCAAAGCCCTCACTACACTGTGTGACTTGGCCGACCGAGAGTTGGTGGTTATCATT 1184

Qy       1007 AGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTC 1066
              | ||||| ||| | ||||| |||||       |||  ||||  |||||  || | ||
Db       1185 GGATGGGCGAAGCATATTCCAGGCTTCTCCACGCTGTCCCTGGCGGACCAGATGAGCCTT 1244

Qy       1067 TTCGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGCGCGCTCAATTGACCAC 1126
              || ||||   |||||||||  | || ||  | ||  |  ||  |  || || ||
Db       1245 CTGCAGAGTGCTTGGATGGAAATTTTGATCCTTGGTGTCGTATACCGGTCTCTTTCGTTT 1304

Qy       1127 CCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTA 1186
              |   | ||  ||| |||   ||| | |  ||||  ||  ||  |||| | |
Db       1305 GAGGATGAACTTGTCTATGCAGACGATTATATAATGGACGAAGACCAGTCCAAATTAGCA 1364

Qy       1187 GAAGGAATTCTGGAAATCTTTGACATGCTCCTG 1219
              |    ||    ||   ||  || || | |
Db       1365 GGCCTTCTTGATCTAAATAATGCTATCCTGCAG 1397


RESULT 22
BJ069238
LOCUS       BJ069238                 610 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ069238 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL052d06 5', mRNA sequence.
ACCESSION   BJ069238
VERSION     BJ069238.1  GI:17494935
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin,I.T. and Kohara
            ,Y.
```

```
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81 559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL052d06"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
                     /note="Vector: pBSRN3; Site_1: Not1; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett  and A.M. Zorn,
                     (Wellcome/CRC Institute). "
BASE COUNT       149 a    144 c    165 g    151 t      1 others
ORIGIN

  Query Match             5.7%;  Score 108.8;  DB 13;  Length 610;
  Best Local Similarity   51.7%;  Pred. No. 9.2e-20;
  Matches  310;  Conservative    0;  Mismatches  272;  Indels   18;  Gaps    2;

Qy        527 GTGTGGAGGGATTAGGATGGGAIAICAAIAIGGAGICIGGIGGIGIGAAGGAICIAAG 586
              || ||  | |||   ||||| || ||||||||||| ||    || ||||| | ||| ||
Db         14 GTGTGTGGAGATGTGGCATCAGGCTATCACTATGGGGTTGCCTCATGTGAGGCATGCAAA 73

Qy        587 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 646
              ||||| || ||  | | ||| |||||  |||  || || |  |||||||| |  ||| |
Db         74 GCCTTCTTCAAGCGCACCATCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATCAA 133

Qy        647 TGTACAATCGATAAAAACCGGCGCAAGAGCTGGCAGGCCTGCCGACTTCGGAAGTCTTAC 706
              |||  |||     || |  || || |||  ||| || || |||||  |   |||||
Db        134 TGTGAAATTACCAAGAGACGTCGAAAGGCCTGTCAAGCTTGCCGCTTCAGAAAGTGCCTT 193

Qy        707 GAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGACATGTGGGTACCGCCTTGTGCCG 766
              |||||  ||| | ||    ||    ||    || |     |     | | |  |
Db        194 CGCGTGGGCATGCTAAAAGAAGGGCTACGCTTGGACGGTGTCAGAGCAGGACGTCAGAAA 253

Qy        767 AGACAGAGAAGTGCCCACCACCAGCTGCACTCTGCCCGCAAGCCCAACAGAAGTCCCCGCC 826
              ||||  ||  | || |   |  | | |                |     ||| ||
Db        254 TACAAGAGGAGCCCAGATGGAGATATCCTC      ------CAGTACTCTCTGTGTGCC 303

Qy        827 CACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTCAGCCCCGAGCAGCTAGTGCTC 886
              || || |  | ||      || |  |   |  | || | | |   ||| ||
Db        304 CAGGCTCAGCAAGCCTCTGCAACAGTGGTCAAGAAGCAAACTCCAGTGAATGCAGTTCTC 363

Qy        887 ACCCTCCTGGAGGCTGAGCCGCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCAGC 946
              | | ||||  || ||     || |   | ||| |      ||      |  | |  |
Db        364 TCTCACCTGCTGGTTGCAGAAGCAGACAAGTTGTTTGCTATGCCAGACCCAGCACTTCCT 423

Qy        947 GAGGCCTCCATG------ATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACAC 1000
              || |       ||      ||||   |||||    |  | || ||    ||  | ||   |
Db        424 GATGGTGATTTGAAATATATGAGCACCCTGTGTGACCTTGCTGATCGTGAAATTGTCATC 483

Qy       1001 ATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTC 1060
              || || |||||||| ||||| || || || ||        ||   ||| |  |||||  ||
Db        484 ATCATTAGCTGGGCAAAGAATATCCCAGGATTCTCCTCCTTTCCCTCTCGGACCAGATG 541

Qy       1061 CGGCTCTTGGAGAGCTGTTGGATGCAGGTGTTAATGATGGGGCTGATGTGGGGCTCAATT 1120
              | ||||||  ||||    ||||||||||| ||| || | ||  |  | |  ||||  ||
Db        542 AGTCTCTTGCAGAGTGTCTGGATGGAGGTTTIATTGCTTGGAGTTGTTTCGGCTCCCTT 601


RESULT 23
BG744031
LOCUS       BG744031                 703 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602722911F1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849155 5',
            mRNA sequence.
ACCESSION   BG744031
VERSION     BG744031.1  GI:14054684
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
            cDNA Library Preparation: Ling Hong/Robin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1688  row: f  column: 04
            High quality sequence stop: 628.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4849155"
                     /clone_lib="NIH_MGC_106"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT       116 a    210 c    245 g    132 t
ORIGIN

  Query Match             5.7%;  Score 107.8;  DB 12;  Length 703;
  Best Local Similarity   48.6%;  Pred. No. 1.9e-19;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    0;  Gaps    0;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | |||||   ||||| | ||    || || || || |||||||| ||    || |||
Db         37 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 96

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db         97 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGAGCATCGAGTACAGCTGTCCG 156

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 694
              ||  | ||  |||||  |||   ||     | ||||||  ||||||||||||||||  |
Db        157 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 216

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGACATGTGGGTAC 754
              |||||        ||||| ||| | |||   ||    ||    || |  ||    ||
Db        217 ACCAAGTGCCTGGGGGTGGGCATGCTCAAGGAGGGAGCGCGCCTGGACGGGTCCGGGGT 276

Qy        755 CGCCTTGTGCCGAGACAGACAAGTGCCGACGAGCAGCTGCTGCTGGCCGGCAAGCCCAAG 814
              | |    | |    || |  | | || | | | |  |       |        ||
Db        277 GGGCGGCAGAACTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTC 336

Qy        815 AGAAGTGGGGGCCAGGGCCCCGAGTGGGGGAGCTGCTGTGGACGCCCTGAGCCCCGAC 874
              |||   ||  ||   |  | |  | | | |  |  |  |  |  |  ||||       |
Db        337 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCA 396

Qy        875 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 934
```

```
              | | |    |     ||  |||  |        |   |         ||       ||||
Db         497 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCA 456

Qy         935 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 994
               | ||          |   | |      ||    ||||      |  |     |||   ||| |
Db         457 GGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATT 516

Qy         995 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 1054
               ||   ||  ||||||||||||||||  || || |||||       |||    ||||  |||
Db         517 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTCTTCAC 576

Qy        1055 CAAGTGCGGCTCTTGCAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGGGC 1114
               ||  ||    |  || |||||   ||||||||||.|| |  || ||||  ||    |||
Db         577 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCAGGCCA 636

Qy        1115 TCAATTG 1121
               |  | ||
Db         637 TGCACTG 644


RESULT 24
BF510841/c
LOCUS       BF510841                 575 bp    mRNA    linear   EST 06-DEC-2000
DEFINITION  UI_H_BI4_apo-e_04_0_UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
            IMAGE:3088206 3', mRNA sequence.
ACCESSION   BF510841
VERSION     BF510841.1  GI:11594139
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html The following repetitive
            elements were found in this cDNA sequence: 1-48,
            -AT_rich#Low_complexity 349-568, -LINE2
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3088206"
                     /clone_lib="NCI_CGAP_Sub8"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
                     is a subtracted library derived from NCI_CGAP_Sub5. The
                     NCI_CGAP Sub8 library had 2.5 million recombinants. A
                     single-stranded DNA preparation of NCI_CGAP_Sub5 was used
                     as a tracer in a subtractive hybridization with a driver
                     comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE
                     clone Ids 2742833-2747415, 3068040-3069191; 25% of the
                     driver population), a pool of clones from NCI_CGAP_Sub4
                     (IMAGE clone Ids 2723592-2729326; 25% of the driver
                     population), NCI_CGAP_Sub6 (pool A1F-A10, IMAGE Ids
                     2728969-2733190; 25% of the driver population); and
                     NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238, 3081864-3084550
                     ; 25% of the driver population). Subtraction was
                     performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To Facilitate Gene Discovery. Genome Research
                     6, 791-806.
                     TAG_LIB-NCI_CGAP_GC4
                     TAG_TISSUE-germ cell
                     TAG_SEQ-AAATC"
BASE COUNT       126 a    131 c    115 g    203 t
ORIGIN

  Query Match             5.7%;  Score 107.4;  DB 12;  Length 575;
  Best Local Similarity   70.2%;  Pred. No. 2.3e-19;
  Matches  144;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         245 AGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCA 304
               || ||||    |  |||  |    |    | |||    |  ||     | |||   ||
Db         447 AGAATTCAAGGAGATGAGAAGGAAACAGGAGGGAAGATTGAAAAGAAGAGTCCAGTGTGT 388

Qy         305 AATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCA 364
                 |||  |      || |       ||  | ||       |||||||||||||||||||||
Db         387 TATGAGGAAAACCCCAAGAGCATGCTGCCTTACAAGACAGTGGTCCATCGCCAGTTATCA 328

Qy         365 CATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACC 424
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         327 CATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACC 268

Qy         425 TTACCTGTAAACAGAGAGACACTGA 449
               |||||||||||||||||||||||||
Db         267 TTACCTGTAAACAGAGAGACACTGA 243


RESULT 25
AK019222
LOCUS       AK019222                 2199 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus 11 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2700080M08:estrogen related receptor,
            alpha, full insert sequence.
ACCESSION   AK019222
VERSION     AK019222.1  GI:12859315
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2700080M08.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system -384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
```

```
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5  (bases 1 to 2199)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,F.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/ Tel:81-45-503-9222
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .2199
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2700080M08"
                     /db_xref="MGD:MGI:1899528"
                     /db_xref="taxon:10090"
                     /clone="2700080M08"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="11 days embryo"
     gene            1. .2199
                     /gene="Esrra"
     misc_feature    1. .2199
                     /gene="Esrra"
                     /note="data source:MGD, source key:MGI:1346831,
                     evidence:ISS
                     estrogen related receptor, alpha"
                     /db_xref="MGD:MGI:1346831"
BASE COUNT      425 a    622 c    722 g    430 t
ORIGIN

  Query Match           5.7%;  Score 107.4;  DB 11;  Length 2199;
  Best Local Similarity   48.7%;  Pred. No. 4.7e-19;
  Matches  322;  Conservative    0;  Mismatches  336;  Indels    3;  Gaps    1;

Qy        515 CACTTCTGGGCTGGTCTGCAGCGATTACGCATGGGATATCACTATGGAGTCTGCTCGTGT 574
              | | |||||   |||||  | |||   || || || || ||||| || ||    || |||
Db        389 CGCCTCTGCCTGGTCTGTGGGCATGTGGCCTCTGGTACCACTACGGTGTGGCATCCTGT 448

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db        449 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGAGCATCAGTGTGCCG 508

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAAGCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 694
              ||  | ||| ||||   |||   ||      | ||||||  ||| ||||||||||  |
Db        509 GCCTCCAATGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGTCAGGCCTGCCGCTTC 568

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTAC 754
              |||||       |||||  ||| | |||   ||    ||    ||    |||    |
Db        569 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGTGTGCGTCTGGA---CGGTGTCCGC 625

Qy        755 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 814
              ||     || ||   | | | |       ||| |   |  |  |  |   |  || |
Db        626 GGCGGAGCGCAGAAGTACAAACGGCGCGCAGAGGAGCCCTTCGCCTTCCGCGGCCGCC 685

Qy        815 AGAAGTGGCGGCCACGCGCCCGAGTGCGGCAGCTGCTGTGGACGCCCTGACCCCGAG 874
              ||  || |   | |  ||  | | |   |  |  |  ||||     ||
Db        686 TTCCCTGCTGGACCTCTGGCAGTAGCTGGAGGACCCAGGAAGACAGCCCCAGTGAACGCT 745

Qy        875 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 934
              | | |  ||    ||  ||| |    |   |      ||    |||
Db        746 CTGGTGTCGCATCTGCTGGTGGTTGAACCTGAGAAGCTGTACGCCATGCCTGACCCAGCA 805

Qy        935 GCGCCCTTCACCGAGGCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 994
              |||       |  | ||   ||   | ||     | |   |    ||  ||| |
Db        806 AGCCCGATGGACACCTCCCCGCTGTGGCACTCTCTGTGACCTTTTGATCCAGAGATA 865

Qy        995 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGCAGCTCAGCCTGTTCGAC 1054
              ||   ||  ||||||||||||||||  || || |||||       ||    |||  |||
Db        866 GTGGTCAACATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCCTCACTGTCACTGTCTGAC 925

Qy       1055 CAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGC 1114
              ||  ||    |  || ||||    |||||||| ||| |  || ||||  ||      |||
Db        926 CAGATGTCAGTACTGCAGAGTGTGTGGATGGAAGTGCTGGTGCTGGGTGTGGCCACGGC 985

Qy       1115 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAG 1174
              ||| |    |  |  |   |||   ||||||   || || ||  |  |    || | |
Db        986 TCACTGCCACTGCAGGATGAGCTGGCCTTTGCTGAGGACCTGGTCTTATATGAAGAGGGG 1045

Qy       1175 G 1175
              |
Db       1046 G 1046


RESULT 26
BG747925
LOCUS       BG747925                676 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602705778F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842402 5',
            mRNA sequence.
ACCESSION   BG747925
VERSION     BG747925.1  GI:14058578
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
```

```
  AUTHORS    NIH_MGC http://mgc.nci.nih.gov/
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1676  row: I  column: 19
             High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "IMAGE:4842402"
                     /clone_lib "NIH_MGC_43"
                     /tissue_type "normal pigmented retinal epithelium"
                     /lab_host "DH10B (phage-resistant)"
                     /note "Organ: eye; Vector: pCFB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: CGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT       113 a    209 c    222 g    132 t
ORIGIN

  Query Match             5.6%;  Score 106.2;  DB 12;  Length 676;
  Best Local Similarity   48.7%;  Pred. No. 5.4e-19;
  Matches  288;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | ||||  ||||| | ||    || || || || |||||||| ||    || ||| 
Db         26 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 85

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || |  || || |||| || || || | ||| || ||    |  || || |  |||| 
Db         86 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGAGCATCGAGTACAGCTGTCCG 145

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 694
              ||  | ||  |||||  |||  ||     | ||||||  ||||||||||||||||  | 
Db        146 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 205

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGCTAC 754
              |||||       ||||| ||| | |||    ||    ||    ||  |      ||  
Db        206 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGT 265

Qy        755 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCCGGCAAGGCCAAG 814
               | |    |  |    || |  |  | || | | | |  |           ||      
Db        266 GGGCGGCAGAAGTACAAGCGGCGGCCGGAGTGGACCCACTGCCCTTCCCGGGCCCCTTC 325

Qy        815 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAG 874
                  |||   ||  ||   |   |   |    | |  |   |  ||||      |    
Db        326 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCCGAAGACAGCAGCCCCAGTCAATGCA 385

Qy        875 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 934
              | | |    |     ||  |||  |       |   |        ||      ||||   
Db        386 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCA 445

Qy        935 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 994
              |  ||         |   | |     ||    ||||     |  |    |||  |||  |
Db        446 GGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATT 505

Qy        995 GTACACATGATCAGCTGGGCCAAGAAGATTCCGGGCTTTGTGGAGCTCAGCCTGTTCGAC 1054
              ||  ,||  ||||||||||||||||  || || |||||      |||    ||||  |||
Db        506 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 565

Qy       1055 CAAGTGCGGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTG 1105
              ||  ||    |  || |||||   ||||||||||| |  || ||||  ||
Db        566 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTG 616


RESULT 27
BI100868
LOCUS       BI100868                 844 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602886184F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5041580
            5', mRNA sequence.
ACCESSION   BI100868
VERSION     BI100868.1  GI:14551761
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11114  row: q  column: 21
            High quality sequence stop: 775.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5041580"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note "Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT       156 a    244 c    276 g    168 t
ORIGIN

  Query Match             5.4%;  Score 102;  DB 14;  Length 844;
  Best Local Similarity   48.7%;  Pred. No. 9.8e-18;
  Matches  307;  Conservative    0;  Mismatches  320;  Indels    3;  Gaps    1;

Qy        546 CGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCA 605
              |||| || ||||| || ||    || ||||| |  || || ||||| || || || | ||
Db         11 CGGGCTACCACTACGGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGGACCA 70

Qy        606 TTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACC 665
              | || ||    |  || || |  ||||| ||  | ||| |||||   |||   ||     
Db         71 TCCAGGGGAGCATCGAGTACAGCTGTCCGGCCTCCAATGAGTGTGAGATCACCAAGCGGA 130

Qy        666 GGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGT 725
              | ||||||  ||| ||||||||||||  |   |||||       ||||| ||| | |||
Db        131 GACGCAAGGCCTGTCAGGCCTGCCGCTTCACCAAGTGCCTGCGGGTGGGCATGCTCAAGG 190

Qy        726 GTGGCTCCCGGAGAGAGAGATGTGGCTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACG 785
              ||    ||    ||  |     |   | | |    |  |    |  |  |  | | | |
Db        191 AGGGTGTGCGTCTGGACCGTGTCCGCGGCGGACGGCAGAAGTACAAACGGCGGCCAGAGG 250

Qy        786 AGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGG 845
              | | |         |  ||         |||   |  ||      |   |  |      
Db        251 TGGACCCTTTGCCTTTCCCGGGCCCCTTCCCTGCTGGACCTCTGGCAGTAGCTGGAGGAC 310

Qy        846 AGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGC 905
                  |  |   | | |  ||| | |   |  |      ||     |  | ||  |||||
```

```
Db        411 CCAGGAAGACAGCCCCAGTGAACGCTCTGGTGTCGCATCTGCTGGTGGTTCAACCTGAGA 470

Qy        906 CGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCTTCACCGAGGCCTCCATGATGATGT 965

Db        471 AGCTGTACGCCATGCCTGACCCAGCAAG---CCCCGATGGACACCTCCCCGCTGTGCCA 527

Qy        966 CCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATCC 1025

Db        528 CTCTCTGTGACCTTTTTGATCGAGAGATACTGGTCACCATCAGCTGGGCCAAGAGCATCC 587

Qy       1026 CCGCCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGC 1085

Db        588 CAGGCTTCTCCTCACTGTCACTGTCTGACCAGATGTCAGTACTGCAGAGTGTGTGGATGC 647

Qy       1086 AGGTGTTAATGATGGGGCTGATGTGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTC 1145

Db        648 AAGTGCTGGTGCTGGGTGTGCCCAGCGCTCACTGCCACTGCAGGATGAGCTGCCTTTC 607

Qy       1146 CTCCAGATCTTGTTCTGGACAGGGATGAGG 1175

Db        608 CTGAGGACCTGGTCCTAGATGAAGAGGGG 637


RESULT 28
BM562894
LOCUS       BM562894                 937 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6588670 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478584
            5', mRNA sequence.
ACCESSION   BM562894
VERSION     BM562894.1  GI:18809360
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 937)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1998  row: h  column: 09
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5478584"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      182 a    286 c    297 g    170 t      2 others
ORIGIN

  Query Match          5.3%;  Score 101.4;  DB 13;  Length 937;
  Best Local Similarity   47.9%;  Pred. No. 1.5e-17;
  Matches  291;  Conservative    0;  Mismatches  316;  Indels    0;  Gaps    0;

Qy        569 TCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATT 628

Db          5 TCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGAGCATCGAGTACAG  64

Qy        629 TGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGC 688

Db         65 TGTCCGGCCTCAACGAGTGTCAGATCACCAACGGGAGACGCAAGGCCTGCAGGCCTGC 124

Qy        689 CGACTTCGGAAGTGTTACGAACTGGCAATGGTGAAGTGTGGCTCCGGGACAGACAGACTT 748

Db        125 CGCTTCACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGCAGTGCGCCTGGACCGCGTC 184

Qy        749 GGGTACCGCCTTGTGCGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAG 808

Db        185 CGGCGTGGGCGGCAGAAGTACAAGCGGCGGCCGCAGCTGCACCCACTGCCCTTCCGGGC 244

Qy        809 GCCAAGAGAAGTGGCGGCCACGGCGCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGC 868

Db        245 CGCTTCCCTGCTGGGCCCTGGCAGTGGTGGAGGCCCCCGGAAGACAGCAGCCCCAGTG 304

Qy        869 CCCCAGCAGCTAGTGCTCACCTCCTTGGAGGCTGAGCGGCCCCATGTGCTGATCAGCGGC 928

Db        305 AATGCACTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGAC 364

Qy        929 CCCAGTGGGCCTTCAGCGAGGCCTCCATGATGATGTCCCTGACCAACTTGGCCGACAAG 988

Db        365 CCCGCAGGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGA 424

Qy        989 GAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTG 1048

Db        425 GAGATTGTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTG 484

Qy       1049 TTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATG 1108

Db        485 TCTGACCAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCC 544

Qy       1109 TGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGG 1168

Db        545 CAGCGCTCACTGCCACTGCAGGATGAGCTGCCTTCGCTGAGGACTTAGTCCTGGATGAA 604

Qy       1169 GATCAGG 1175

Db        605 GAGGGGG 611


RESULT 29
BJ058831
LOCUS       BJ058831                 558 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ058831 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL060113 5', mRNA sequence.
ACCESSION   BJ058831
VERSION     BJ058831.1  GI:17492544
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Xenopus laevis"
```

```
                     /db_xref "taxon:8355"
                     /clone "XL060113"
                     /clone_lib "NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type "whole embryo"
                     /dev_stage "stage 25"
                     /note "Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett  and A.M. Zorn,
                     (Wellcome/CRC Institute). "
BASE COUNT      142 a    134 c    146 g    136 t
ORIGIN

  Query Match             5.4%;  Score 100.4;  DB 14;  Length 558;
  Best Local Similarity   51.6%;  Pred. No. 2.3e-17;
  Matches  292;  Conservative    0;  Mismatches  256;  Indels   18;  Gaps    2;

Qy        542 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGA 601
              ||||| || |||||||||||| ||    || ||||| | ||| || |||||| || ||  |
Db          1 GCATCAGGCTATCACTATGGGGTTGCCTCATGTGACGCATGCAAAGCCTTCTTCAAGCGC 60

Qy        602 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAA 661
              | ||| ||||| ||| || || | ||||||| | ||| | ||| ||| |    ||
Db         61 ACCATCCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATGAATGTGAAATTACCAAG 120

Qy        662 AACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTG 721
              |  || || |||  ||| || || ||||| |   ||||        ||||| ||| | |
Db        121 AGACGTCGAAAGGCCTGTCAAGCTTGCCGCTTCACAAAGTGCCTTCGCGTGGGCATGCTA 180

Qy        722 AACTGTGGCTCCCCGCAGCAGCAGATGTCGCTACCGCCTTGTGCCGAGACAGACAGTGCC 781
              ||    ||    ||    ||  |     |     | | | |       |||| || | |
Db        181 AAAGAAGGGGTAGCTTGGACCGTGTCAGAGGAGGACGTCAGAAATACAAGAGGAGGCCA 240

Qy        782 GACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTG 841
              || |  | | | |                 ||| || || || |  | ||
Db        241 GATGGAGATATCCTC----------CAGTACTCGTCTGGTGGGCAGGCTCAGCAAGCC 288

Qy        842 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 901
              |     || |  |   |  | || | | |   ||| || | | ||||  || |
Db        289 TCTGCAACAGTGGTCAAGAAGCAAACTTCCAGTGAATGCAGTTCTCTCTCACCTGCTGGTT 348

Qy        902 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATG -- 958
              |     ||  |   | || |       ||      |  |  |  | || |       ||
Db        349 GCAGAACCAGACAAGTTGTTTGCTATGCCAGACCCAGCACTTCCTGATGGTGATTTGAAA 408

Qy        959    ATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
                 ||||  |||||   | | || ||    || | || |||  || |||||||||
Db        409 TATATGAGCACCCTGTGTGACCTTGCTGATCGTGAAATTGTCATCATCATTAGCTGGGCA 468

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||||| || || || ||        ||   ||| |  ||||| || | |||||| ||||
Db        469 AAGAATATCCCAGGATTCTCCTCCCTTTCCCTCTCGGACCAGATGAGTCTCTTGCAGAGT 528

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGG 1101
              |||||||||||| ||| || | ||
Db        529 GTCTGGATGGAGGTTTTATTGCTTGG 554


RESULT 40
BE546976
LOCUS       BE546976                908 bp    mRNA    linear   EST 09-AUG-2000
DEFINITION  601071950F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3457885 5',
            mRNA sequence.
ACCESSION   BE546976
VERSION     BE546976.1  GI:9775621
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 908)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8448  row: d  column: 14
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3457885"
                     /clone_lib "NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      172 a    251 c    334 g    151 t
ORIGIN

  Query Match             5.3%;  Score 99.8;  DB 10;  Length 908;
  Best Local Similarity   48.3%;  Pred. No. 4.4e-17;
  Matches  278;  Conservative    0;  Mismatches  297;  Indels    0;  Gaps    0;

Qy        542 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAACA 601
              || || || || |||||||| ||    || ||||| |  || || ||||| || || ||
Db         11 GCCTCCGGCTACCACTATGGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGG 70

Qy        602 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAA 661
              | ||| || ||    |  || || | ||||| ||  | ||  |||||   |||  ||
Db         71 ACCATCCAGGGGAGCATCGAGTACAGCTGTCCGGCCTCCAACGAGTGTGAGATCACCAAG 130

Qy        662 AACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTG 721
              | ||||||  |||||||||||||||  |    |||||          ||||| ||| | |
Db        131 CGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAAGTGCCTGCGGGTGGGCATGCTC 190

Qy        722 AAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCCGAGACAGACAGAAGTGCC 781
              |||  ||    ||   || |    ||    | |  |  |  | | | | | | | | |
Db        191 AAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGTGGGCGGCAGAAGTACAAGCGGCGGCCG 250

Qy        782 GACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTG 841
              || | | | |  |       |       ||         |||   ||  ||   |  |
Db        251 GAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCCCTGCTGGGCCCCTGGCAGTCGCTGGA 310

Qy        842 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 901
              |    | | |   |  ||||        | | | |    |    ||  |||  |
Db        311 GGCCCCCGGAAGACAGCAGCCCCACGAATGCACTGGTGTCTCATCTGCTGGTGGTTGAG 370

Qy        902 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATG 961
                  |    |          ||       ||||  |  ||          |  | |       ||
Db        371 CCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCCCTGATGGGCACCTCCCAGCCGTG 430

Qy        962 ATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAG 1021
                ||||      |  |    |||   ||| | ||   ||  ||||||||||||||||
Db        431 GCTACCCTCTGTGACCTCTTTGACCGAGAGATTGTGGTCACCATCAGCTGGGCCAAGAGC 490

Qy       1022 ATCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGG 1081
              || || ||||       |||    |||| |||||  ||     |  || |||||   |||
Db        491 ATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAGATGTCAGTACTGCAGAGCGTGTGG 550

Qy       1082 ATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTC 1116
              |||||||||  |  || ||||  ||     ||| | |
Db        551 ATGGAGGTGCTGGTGCTGGGTGTGGCCAGGCTCAC 585
```

```
RESULT 31
AU169990
LOCUS       AU169990                800 bp    mRNA    linear   EST 29-JAN-2001
DEFINITION  AU169990 Ol-br-ad cDNA Oryzias latipes cDNA clone br5464, mRNA
            sequence.
ACCESSION   AU169990
VERSION     AU169990.1  GI:12592059
KEYWORDS    EST.
SOURCE      Japanese medaka.
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Mita,K., Ishikawa,Y. and Yamauchi,M.
  TITLE     Establishment of cDNA database of medaka, Oryzias latipes
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mita K
            Genome Research Group
            National Institute of Radiological Sciences
            Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
            Email: kmita@nirs.go.jp
            method:uni-directional sequence direction:sequenced from T3 primer
            (5' -> 3').
FEATURES             Location/Qualifiers
     source          1..800
                     /organism="Oryzias latipes"
                     /strain="HNI"
                     /db_xref="taxon:8090"
                     /clone="br5464"
                     /clone_lib="Ol-br-ad cDNA"
                     /sex="female/male mixed"
                     /tissue_type="brain"
                     /dev_stage="adult"
BASE COUNT      198 a    212 c    238 g    152 t
ORIGIN

  Query Match             5.2%;  Score 99.6;  DB 9;  Length 800;
  Best Local Similarity   65.0%;  Pred. No. 4.7e-17;
  Matches  147;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        519 TCTGGGCTGTCTGCAGGGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 578
              ||||| || || | |||  || || || ||||||||||||||  ||| |||||||
Db        568 TCTGGCTCGTATGTGGGGATGTGGCGTCCGGGTATCACTATGGACTGGGGGCATGTGAAG 627

Qy        579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638
              || || || || || || |||| |||||||||  | |  || || |  ||||||||
Db        628 CCTGCAAAGCTTTGTTCAAGAGAACCATTCAAGGTAACATGGAGTACAGCTGTCCAGCAT 687

Qy        639 CAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTGGA 698
              ||||  ||||  ||||  |||| ||  | |||  ||||| || |||||  ||   |
Db        688 CAAACGAGTGCGAAATCACCAAAACACGCAGAAAGGGTGCCAAGCATGCCGCTTACCA 747

Qy        699 AGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAG 744
              ||||  | |||| || ||| | |||  ||  ||  ||| ||
Db        748 AGTGCCTCAAACTAGCCATGCTCAAGGAGGGAGTGCGTCTAGATAG 793


RESULT 32
AL527012
LOCUS       AL527012                1004 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL527012 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC017YG10 5
            prime, mRNA sequence.
ACCESSION   AL527012
VERSION     AL527012.1  GI:12790505
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web . www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..1004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DC017YG10"
                     /clone_lib="LTI_NFL003_NBC3"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 401 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      163 a    324 c    350 g    164 t      3 others
ORIGIN

  Query Match             5.2%;  Score 99.6;  DB 9;  Length 1004;
  Best Local Similarity   49.0%;  Pred. No. 5.8e-17;
  Matches  290;  Conservative    1;  Mismatches  300;  Indels    1;  Gaps    1;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | |||||   |||||  | ||    || || || || |||||||| ||    || |||
Db        380 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 439

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || |  || || ||||| || || || | ||| || ||    |  || || |  ||||||
Db        440 GAGGCCTGCAAAGCCTTCTTCAAGAGCACCATCCAGGGGAGCATCGAGTACAGCTGTCCA 499

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTT 694
              ||  | ||  |||||   |||   ||      | ||||||  ||||||||||||||  |
Db        500 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 559

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGTAC 754
                |||||      ||||| ||| | |||   ||     ||    || |     ||
Db        560 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGG 619

Qy        755 CGCCTTGTGGGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 814
              |    |      |||      |||| |   |  |||||  |   |  ||||
Db        620 TGGGCGGCAGAAGTACAAGGGGGGCGCGGAGGTGGGACCCACTGCCCTCCCGGGGCCCT 679

Qy        815 AGAAGTGGGGGCCACGCGCCCCGAGTGCGGGAGCTGCTGTGGACGCCTGAGCCCGAG 874
                  |||| |  ||      ||  ||  |      |    |           |
Db        680 TCCCTGCTGGGCCCCTGGAGTCGCTGGAGGCCCCGGAAGACAGCAGCCCCAGTGAATG 739

Qy        875 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCCATGTGCTGATCAGCCGCCCAGT 934
              ||      |||| |  |  |  |      | |          ||    |||   ||
Db        740 CACTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCG 799

Qy        935 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 994
              |||||     |  | |    ||    ||||    |  |    |||   ||| |
Db        800 CAGCCCGATGGGCACTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATT 859

Qy        995 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGACTCAGCCTGTTCGAC 1054
              ||   || ||||||||||||||||  || |  |||||       |||    ||||  |||
Db        860 GTGGTCACCATCAGCTGGGCCAAGAGCATCCAAGGCTTCTCATCGCTGTCGCTGTCTGAC 919

Qy       1055 CAAGTGCGGCTCTTGGAGAGC-TGTTGGATGGAGGTGTTAATGATGGGGCTG 1105
              ||  ||    |  || ||||:| || ||||| ||||||  |  || ||| | ||
Db        920 CAGATGTCAGTACTGCAGACGGTGGTGGATTGAGGTGCTGGTGCTGGTGGTG 971
```

```
RESULT 33
BG480960
LOCUS       BG480960                775 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602529118F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4652501 5',
            mRNA sequence.
ACCESSION   BG480960
VERSION     BG480960.1  GI:13413239
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1437  row: d  column: 06
            High quality sequence stop: 663.
FEATURES             Location/Qualifiers
     source          1..775
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:4652501"
                     /clone_lib "NIH_MGC_21"
                     /tissue_type "choriocarcinoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G)  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      141 a    216 c    282 g    146 t
ORIGIN

  Query Match             5.1%;  Score 96.6;  DB 12;  Length 775;
  Best Local Similarity   51.1%;  Pred. No. 4.3e-16;
  Matches  403;  Conservative    0;  Mismatches  284;  Indels    6;  Gaps    3;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | |||||   |||||  | ||    || || || || ||||||| ||    || |||
Db         83 CGCCTCTGCCTGGTCTGTGGGCACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 142

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || |  || || ||||| || || || | ||| || |||   |  || || |  |||||
Db        143 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGA-GCATCGAGTACAGCTGTCCG 201

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAACCGGCCCAAGAGCTGCCAGGCCTGCCGACTT 694
              ||  | ||  |||||   |||   ||      | ||||||  ||||||||||||||||  |
Db        202 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACCCAAGGCCTGCCAGGCCTGCCGCTTC 261

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAG--TGTGGCTCCCGGAGAGACAGATGTGGGT 752
              |||||       |||||  ||| | |||   |  |  | |   ||  | |  | ||
Db        262 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCGGGGTG 321

Qy        753 ACCGCCTTGTGCCGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCA 812
                || |    |  | | | |   | |  || |  |||| ||   | |||     |
Db        322 GGCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCC 381

Qy        813 AGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCG 872
                  |  |      |  | |||  |   | |      | | |  |||   |
Db        382 CTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCAC 441

Qy        873 AGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCA 932
              |  |       ||     |  | ||| |||||  || | |||   ||     || | |
Db        442 TGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAG 501

Qy        933 GTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGT 992
              |  |          |   |  || ||   |  ||||     |  |   |||  |||
Db        502 GCCCTGATGGGCACCTCCCAGCCGTGGCTA---CCCTCTGTGACCTCTTTGACCGAGAGA 558

Qy        993 TGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCG 1052
              | ||   ||  |||||||||||||||||  || || ||||      |||    |||| |
Db        559 TTGTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTG 618

Qy       1053 ACCAACTCCGGGCTCTTCGCAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTG 1105
              ||||  ||    |  || |||||   ||||||||||||| |  || ||||  ||
Db        619 ACCAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTG 671


RESULT 34
BG033384
LOCUS       BG033384                979 bp    mRNA    linear   EST 24 JAN 2001
DEFINITION  602301556F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4403197 5',
            mRNA sequence.
ACCESSION   BG033384
VERSION     BG033384.1  GI:12425627
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10112  row: p  column: 14
            High quality sequence stop: 681.
FEATURES             Location/Qualifiers
     source          1..979
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4403197"
                     /clone_lib="NIH_MGC_87"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: breast; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.384 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      150 a    291 c    353 g    184 t      1 others
ORIGIN

  Query Match             5.1%;  Score 96.4;  DB 12;  Length 979;
  Best Local Similarity   49.0%;  Pred. No. 4.4e-16;
  Matches  285;  Conservative    0;  Mismatches  296;  Indels    1;  Gaps    1;

Qy        542 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGA 601
              || || || || |||||||| ||    || ||||| |  || || ||||| || || ||
Db          4 GCCTCCGGCTACCACTATGGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGG 63

Qy        602 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAA 661
              | ||| || |||   |  || || |  |||||||  | ||  |||||   |||   ||
Db         64 ACCATCCAGGGAGC-ATCGAGTACAGCTGTCCAGCCTCCAACGAGTGTGAGATCACCAAG 122
```

```
Qy        662 AACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGCAATGCTG 721
              | ||||||  ||||||||||||||||  |    |||||          ||||| ||| |
Db        123 CGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAAGTGCCTGCGGGTGGGCATGCTC 182

Qy        722 AAGTGTGGCTCCGGCAGACAGACATGTGGGTACGGCCTTGTGCGGAAGACAGAGAAGTGCC 781
              |||   ||    ||    ||  |     ||    | |     |  |    || |   | |
Db        183 AAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGTGGGCGGCAGAAGTACAAGCGGCGGCCG 242

Qy        782 GACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTG 841
              || | | | |  |       |       ||        |||   ||  ||   |   |
Db        243 GAGGTGGACCCACTGCCCTTCCCGGGCTCCTTCCCTGCTGGGCCCCTGGCAGTCGCTGGA 302

Qy        842 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 901
              |   | | |   |  ||||        |  | | |    |    ||  |||  |
Db        303 GGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACTGGTGTCTCATCTCCTCGTGCTTGAG 362

Qy        902 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTCCGCCCTTCACCGAGGCCTCCATGATG 961
              |   |       ||      ||||   | ||       |   | |     ||
Db        363 CCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCCCTGATGGGCACCTCCCAGCCCTG 422

Qy        962 ATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAG 1021
              ||||     |  |    |||  ||| | ||   ||  ||||||||||||||||
Db        423 GCTACCCTCTGTGACCTCTTTGACCGAGAGATTGTGGTCACCATCAGCTGGGCCAAGAGC 482

Qy       1022 ATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGC 1081
              || || |||||      |||   ||||  |||||  ||    |  || |||||   |||
Db        483 ATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAGATGTCAGTACTCCAGAGCGTGTGC 542

Qy       1082 ATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGAC 1123
              |||||||||  |  || ||||  ||      |||||| |  ||
Db        543 ATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGCTCACTGCAC 584
```

```
RESULT 35
BI738331
LOCUS       BI738331                818 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603861587F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5368892 5',
            mRNA sequence.
ACCESSION   BI738331
VERSION     BI738331.1  GI:15715444
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11938  row: m  column: 21
            High quality sequence stop: 817.
FEATURES             Location/Qualifiers
     source          1..818
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5368892"
                     /clone_lib "NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
```

```
BASE COUNT      185 a    245 c    230 g    158 t
ORIGIN

  Query Match           5.0%;  Score 95;  DB 13;  Length 818;
  Best Local Similarity   65.1%;  Pred. No. 9.9e-16;
  Matches  140;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | | |||   ||  ||| | ||    || || || || |||||  |||||    || ||
Db        303 CCCCTGTGCCTCGTGTGCGGGGACATTGCCTCTGGCTACCACTACGCAGTGGCCTCCTGC 362

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || |  || ||||| || || || |||| |||||||||  | |  || || |  || ||
Db        363 GAGGCTTGCAAGGCGTTCTTCAAGAGAAGCATTCAAGGCAACATCGAGTACAACTGCCCG 422

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 694
              || || ||| | |||   |||   |||   ||||||||  ||| ||||||||||| |
Db        423 GCCACCAATGAATGTGAGATCACCAAACGGAGGCGCAAGTCCTGTCAGGCCTGCCGATTC 482

Qy        695 CGGAAGTGTTACGAAGTGGGAAIGGTGAAGTGTGG 729
              ||| ||   | |||||| ||| |||||   ||
Db        483 ATGAAATGCCTCAAAGTGGGGATGCTGAAGGAAGG 517
```

```
RESULT 36
BF717864
LOCUS       BF717864                402 bp    mRNA    linear   EST 07-JUN-2001
DEFINITION  fd52h03.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
            IMAGE:3733397 5' similar to TR:O13012 O13012 ESTROGEN RECEPTOR. ;,
            mRNA sequence.
ACCESSION   BF717864
VERSION     BF717864.1  GI:12016963
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,K., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 401.
FEATURES             Location/Qualifiers
     source          1..402
                     /organism "Danio rerio"
                     /db_xref "taxon:7955"
                     /clone="IMAGE:3733397"
                     /clone_lib "Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type "26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
```

/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik,Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 112 a 113 c 103 g 74 t
ORIGIN

Query Match 4.9%; Score 93; DB 12; Length 402;
Best Local Similarity 71.1%; Pred. No. 2.5e-15;
Matches 123; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy       1398 TTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGC 1457
             || ||| | || | ||||| |||| ||| ||||| || || ||| | || || ||||
Db          1 TTTCCAGAACGGGTCTGTCCTCACAGCGGCAGTCCATCCGGCTCGCTCATCTGCTAATGC 60

Qy       1458 TCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGT 1517
             | || || ||| || | ||     || ||||| ||||| || ||||||   |||||||||
Db         61 TGCTCTCACACATCCGACACCTCAGCAACAAAGGCATCGAGCATCTGTCAAACATGAAGA 120

Qy       1518 GCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCAC 1570
             | ||||| ||||| |   |||| || || ||||| |||||||||| | ||  ||
Db        121 GGAAAAACGTGGTGCTGCTGTACGATCTTCTGCTAGAGATGCTGGACGCAAAC 173
```

RESULT 37
BM957249
LOCUS BM957249 669 bp mRNA linear EST 18-MAR-2002
DEFINITION fy75h10.y1 Suqano SJD adult male Danio rerio cDNA clone 5777131 5' similar to TR:Q90272 Q90272 RETINOIC ACID RECEPTOR ALPHA 2.B. , mRNA sequence.
ACCESSION BM957249
VERSION BM957249.1 GI:19540679
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopteryqii; Neopteryqii; Teleostei; Ostariophysi; Cypriniformes ; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 669)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Other ESTs: fy75h10.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
Library constructed by Dr. Sumio Suqano and Dr. Koichi Kawakami DNA Sequencinq by: Washinqton University Genome Sequencing Center Clone distribution information can be found throuqh the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.qov
High quality sequence stop: 526.

FEATURES Location/Qualifiers
source 1. .669
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5777131"
/clone_lib="Suqano SJD adult male"
/sex="male"
/tissue_type="whole body"
/dev_staqe="adult"
/lab_host="DH10B (phaqe resistant)"
/note "Vector: pME18S-FL3; Site_1: DraIII (CACCATGTG); Site_2: DraIII (CACTGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Suqano (University of Tokyo Institute of Medical Science). Custom primers for sequencinq: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA. "

BASE COUNT 182 a 174 c 183 g 130 t
ORIGIN

Query Match 4.8%; Score 91.8; DB 14; Length 669;
Best Local Similarity 60.7%; Pred. No. 7.3e-15;
Matches 150; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy        477 CCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCG 536
              |||||||| | |   ||| | ||  ||| |   | ||| ||||||| |
Db        220 CCAGCCCTCCTTCACCGCCACCTCCACCACGGGTCTACAAGCCCTGTTTTGTCTGCCAGG 279

Qy        537 ATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTA 596
              |  | | ||||| || |||||||||||| |  | ||||||||  |||||||  |||||
Db        280 ACAAGTCTTCGGGCTACCACTATGGAGTCAGCGCATGTGAAGGGTGTAAGGGTTTTTCC 339

Qy        597 AAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCG 656
              ||||||||||||     | |  |  || |  || |         ||   ||||  |||
Db        340 GAAGAAGCATTCAGAAGAACATGGTGTACACCTGCCACAGAGAGAAGAGCTGTATCATCA 399

Qy        657 ATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAA 716
              | ||   |   || ||  ||||||||  |||||||||  | ||||||    ||||||||
Db        400 ACAAGGTCACCCGAAACCGCTGCCAGTACTGCCGACTGCAGAAGTGCCTGGAAGTGGGCA 459

Qy        717 TGGTGAA 723
              ||   ||
Db        460 TGTCTAA 466
```

RESULT 38
BB711891
LOCUS BB711891 434 bp mRNA linear EST 11-OCT-2001
DEFINITION BB711891 RIKEN full-length enriched, 2 cells eqq Mus musculus cDNA clone B020030L17 3', mRNA sequence.
ACCESSION BB711891
VERSION BB711891.1 GI:16065060
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 434)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanaqaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T. , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinaqawa,A., Shiraki,T., Soqabe,Y., Suzuki,H., Taqawa ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T. , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full length cDNAs (Akimura,T., et al.

```
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro cho, Tsurumi ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             Wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia. real time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="B020030L17"
                     /clone_lib="RIKEN full-length enriched, 2 cells egg"
                     /tissue_type="egg"
                     /dev_stage="2 cells"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GACAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGACTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector, a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC 1."
BASE COUNT      116 a     99 c    107 g    112 t
ORIGIN

  Query Match             4.8%;  Score 91.2;  DB 10;  Length 434;
  Best Local Similarity   66.0%;  Pred. No. 8.6e-15;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        519 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 578
              |||||   ||||||   ||| | || || |||| ||| ||||| |||| | | |||| |
Db        219 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGGTGCTGACGTGTGGAA 278

Qy        579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638
              | ||||| | ||| |||||||||    | |||||||  | |||| |||||  | |  |
Db        279 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGGTGGAA 338

Qy        639 CAAATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 698
              |||| | || |  || |||||||  ||  | || | ||| |  || |||||   ||| |
Db        339 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 398

Qy        699 AGTGTTACGAAGTGGGAATG 718
              | |||    |||  ||||||
Db        399 AATGTCTTCAAGCTGGAATG 418


RESULT 39
BG687711
LOCUS       BG687711                 521 bp    mRNA    linear   EST 02-MAY-2001
DEFINITION  334912 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BG687711
VERSION     BG687711.1  GI:13929512
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and
            Wells,K.D.
  TITLE     Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@anri.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 33  row: E  column: 5
            Seq primer: ATITAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="BARC 5BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."
BASE COUNT      162 a    111 c    114 g    134 t
ORIGIN

  Query Match             4.8%;  Score 91.2;  DB 12;  Length 521;
  Best Local Similarity   66.0%;  Pred. No. 9.5e-15;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        519 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 578
              |||||   ||||||   ||| | || || |||| ||| ||||| |||| | | |||| |
Db        225 TCTGCCTAGTCTGCTCTGATGAAGCTTCAGGATGTCATTATGGGGTCTTGACTTGTGGAT 284

Qy        579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638
              | ||||| |  || || ||||||     |  |||||||  |  ||||| |||| | |  |
Db        285 GCTGTAAAGTATTCTTCAAAAGAGCAGTGGAAGGACAGCACAATTATCTTTGTGCTGGAA 344

Qy        639 CAAATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 698
              |||| | ||||  || |||||||  ||  | || | |||||  || |||||   | | |
Db        345 GAAATGATTGTATCATTGATAAAATTCGAAGAAAAAACTGCCCAGCATGCCGCTATAGAA 404

Qy        699 AGTGTTACGAAGTGGGAATG 718
              | ||     |||  ||||||
Db        405 AATGCCTTCAAGCTGGAATG 424
```

```
RESULT 40
BB196002
LOCUS       BB196002                 698 bp    mRNA    linear   EST 31-AUG 2001
DEFINITION  BB196002 RIKEN full length enriched, adult male spinal cord Mus
            musculus cDNA clone A330092E23 3' similar to X13358 Mouse mRNA for
            glucocorticoid receptor form A, mRNA sequence.
ACCESSION   BB196002
VERSION     BB196002.2  GI:15409595
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jun 30, 2000 this sequence version replaced gi:8860955.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81 45 503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            Waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system- 384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer based methods for the mouse full length cDNA
            encyclopedia: real time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
            Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, 172-186 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..698
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone="A330092E23"
                     /clone_lib "RIKEN full-length enriched, adult male spinal
                     cord"
                     /sex="male"
                     /tissue_type "spinal cord"
                     /dev_stage "adult"
                     /lab_host "DH10B"
                     /note "Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot - 10.0 and subtraction to Rot - 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC 1. "
BASE COUNT      193 a    149 c    173 g    183 t
ORIGIN

  Query Match             4.8%;  Score 91.2;  DB 10;  Length 698;
  Best Local Similarity   66.0%;  Pred. No. 1.1e-14;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         519 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 578
               ||||| || ||| |||| | || ||||||| || ||||| || | |||||| |
Db         485 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGGTGCTGACGTGTGGAA 544

Qy         579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638
               | ||||| | ||| |||||||||     | |||||||  | ||||  |||||  | |  |
Db         545 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 604

Qy         639 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 698
               |||| | || |  || |||||||  ||  | || | ||| |  || |||||   ||| |
Db         605 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 664

Qy         699 AGTGTTACGAAGTGGGAATG 718
               | |||    |||  ||||||
Db         665 AATGTCTTCAAGCTGGAATG 684


Search completed: June 16, 2003, 03:16:11
Job time : 2891.85 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 13:13:43 ; Search time 26.0717 Seconds
(without alignments)
598.126 Million cell updates/sec

Title: US-08-826-361A-25
Perfect score: 2805
Sequence: 1 MDIKNSPSSLNSPSSYNCSQ.....ECSPAEDSKSKEGSQNPQSQ 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2805	100.0	548	4	US-09-139-617-1	Sequence 1, Appli
2	2805	100.0	548	4	US-09-561-741A-1	Sequence 1, Appli
3	2554	91.1	485	2	US-08-836-620A-2	Sequence 2, Appli
4	2291	81.7	485	2	US-08-836-620A-2	Sequence 2, Appli
5	2286	81.5	484	2	US-08-836-620A-13	Sequence 13, Appl
6	2267	80.8	485	2	US-08-836-620A-5	Sequence 5, Appli
7	2262	80.6	484	2	US-08-836-620A-14	Sequence 14, Appl
8	2017	71.9	384	2	US-08-836-620A-15	Sequence 15, Appl
9	1233.5	44.0	595	4	US-08-764-870-12	Sequence 12, Appl
10	1233.5	44.0	595	4	US-08-980-115-12	Sequence 12, Appl
11	1224.5	43.7	595	4	US-09-041-886-35	Sequence 35, Appl
12	1224.5	43.7	595	4	US-08-458-908-2	Sequence 2, Appli
13	1222.5	43.6	591	2	US-08-836-620A-17	Sequence 17, Appl
14	1214.5	43.3	596	2	US-08-836-620A-16	Sequence 16, Appl
15	1160	41.4	410	6	5223606-5	Patent No. 5223606
16	1126.5	40.2	229	4	US-09-249-645-1	Sequence 1, Appli
17	1064	37.9	228	4	US-09-249-645-2	Sequence 2, Appli
18	1055	37.6	226	2	US-08-836-620A-7	Sequence 7, Appli
19	810	28.9	773	3	US-08-564-264-1	Sequence 1, Appli
20	762	27.2	651	4	US-08-693-940-3	Sequence 3, Appli
21	762	27.2	651	4	US-09-566-660-3	Sequence 3, Appli
22	668.5	23.8	243	2	US-08-836-620A-9	Sequence 9, Appli
23	666	23.7	243	2	US-08-836-620A-8	Sequence 8, Appli
24	660.5	23.5	243	2	US-08-836-620A-10	Sequence 10, Appl
25	648	23.1	435	4	US-09-040-508-2	Sequence 2, Appli
26	648	23.1	435	4	US-09-500-654-2	Sequence 2, Appli
27	648	23.1	458	3	US-09-141-000-4	Sequence 4, Appli
28	630	22.5	500	3	US-09-141-000-2	Sequence 2, Appli
29	627.5	22.4	433	2	US-08-836-620A-19	Sequence 19, Appl
30	588.5	21.0	518	2	US-08-836-620A-18	Sequence 18, Appl
31	576	20.5	418	3	US-09-141-000-6	Sequence 6, Appli
32	460.5	16.4	525	4	US-08-764-870-7	Sequence 7, Appli
33	460.5	16.4	525	4	US-08-980-115-7	Sequence 7, Appli
34	459	16.4	446	1	US-07-952-800-4	Sequence 4, Appli
35	458.5	16.3	533	1	US-07-952-800-2	Sequence 2, Appli
36	452	16.1	463	1	US-08-336-408B-6	Sequence 6, Appli
37	452	16.1	463	5	PCT-US91-00399-6	Sequence 6, Appli
38	451	16.1	363	6	5223606-6	Patent No. 5223606
39	451	16.1	933	4	US-08-764-870-14	Sequence 14, Appl
40	451	16.1	933	4	US-08-980-115-14	Sequence 14, Appl
41	447.5	16.0	984	4	US-08-764-870-15	Sequence 15, Appl
42	447.5	16.0	984	4	US-08-980-115-15	Sequence 15, Appl
43	442.5	15.8	467	1	US-08-336-408B-4	Sequence 4, Appli
44	442.5	15.8	467	5	PCT-US91-00399-4	Sequence 4, Appli
45	438.5	15.6	777	4	US-08-764-870-13	Sequence 13, Appl
46	438.5	15.6	777	4	US-08-980-115-13	Sequence 13, Appl
47	438	15.6	462	1	US-08-336-408B-2	Sequence 2, Appli
48	438	15.6	462	4	US-08-764-870-6	Sequence 6, Appli
49	438	15.6	462	4	US-08-980-115-6	Sequence 6, Appli
50	438	15.6	462	5	PCT-US91-00399-2	Sequence 2, Appli
51	437.5	15.6	452	4	US-08-764-870-16	Sequence 16, Appl
52	437.5	15.6	452	4	US-08-980-115-16	Sequence 16, Appl
53	437.5	15.6	918	4	US-09-041-886-11	Sequence 11, Appl
54	426.5	15.2	1070	4	US-09-091-042A-2	Sequence 2, Appli
55	420.5	15.0	356	6	5223606-7	Patent No. 5223606
56	417.5	14.9	795	1	US-07-716-827C-5	Sequence 5, Appli
57	397.5	14.2	558	1	US-08-333-358-12	Sequence 12, Appl
58	397.5	14.2	558	1	US-08-463-694-12	Sequence 12, Appl
59	397.5	14.2	558	1	US-08-694-501-12	Sequence 12, Appl
60	384	13.7	560	2	US-09-132-619-10	Sequence 10, Appl
61	384	13.7	560	3	US-09-282-803B-10	Sequence 10, Appl
62	384	13.7	560	4	US-09-510-654-10	Sequence 10, Appl
63	383	13.7	495	2	US-09-132-619-2	Sequence 2, Appli
64	383	13.7	495	3	US-09-282-803B-2	Sequence 2, Appli
65	383	13.7	495	4	US-09-510-654-2	Sequence 2, Appli
66	383	13.7	541	2	US-09-132-619-6	Sequence 6, Appli
67	383	13.7	541	3	US-09-282-803B-6	Sequence 6, Appli
68	383	13.7	541	4	US-09-510-654-6	Sequence 6, Appli
69	381	13.6	66	2	US-08-836-620A-11	Sequence 11, Appl
70	381	13.6	674	4	US-08-653-648A-14	Sequence 14, Appl
71	380.5	13.6	500	2	US-09-132-619-8	Sequence 8, Appli
72	380.5	13.6	500	3	US-09-282-803B-8	Sequence 8, Appli
73	380.5	13.6	500	4	US-09-510-654-8	Sequence 8, Appli
74	380	13.5	66	2	US-08-836-620A-12	Sequence 12, Appl
75	374.5	13.3	513	1	US-08-464-266-2	Sequence 2, Appli
76	374.5	13.3	513	1	US-08-464-272-2	Sequence 2, Appli
77	374.5	13.3	513	4	US-08-464-514-2	Sequence 2, Appli
78	374.5	13.3	513	4	US-08-486-403-2	Sequence 2, Appli
79	372	13.2	455	3	US-09-038-217A-18	Sequence 18, Appl
80	370.5	13.2	454	4	US-08-764-870-5	Sequence 5, Appli
81	370.5	13.2	454	4	US-08-980-115-5	Sequence 5, Appli
82	370.5	13.2	494	6	5260432-2	Patent No. 5260432
83	369.5	13.2	414	1	US-08-255-471-9	Sequence 9, Appli
84	367	13.1	462	2	US-08-592-383-2	Sequence 2, Appli
85	367	13.1	462	2	US-08-095-728B-4	Sequence 4, Appli
86	367	13.1	462	5	PCT-US92-02320A-4	Sequence 4, Appli
87	363	12.9	448	6	5223606-2	Patent No. 5223606
88	362	12.9	567	4	US-08-927-219-131	Sequence 131, App
89	362	12.9	465	4	US-08-927-219-79	Sequence 79, Appl
90	361.5	12.9	368	6	5223606-3	Patent No. 5223606
91	361.5	12.9	516	4	US-08-927-219-140	Sequence 140, App
92	360	12.8	440	1	US-08-333-358-8	Sequence 8, Appli
93	360	12.8	440	1	US-08-463-694-8	Sequence 8, Appli
94	360	12.8	440	1	US-08-694-501-8	Sequence 8, Appli
95	360	12.8	447	1	US-08-373-935-1	Sequence 1, Appli
96	359	12.8	462	6	5171671-2	Patent No. 5171671
97	358.5	12.8	403	2	US-08-592-383-4	Sequence 4, Appli
98	356.5	12.7	461	1	US-08-330-518-2	Sequence 2, Appli
99	356.5	12.7	461	1	US-08-330-283-2	Sequence 2, Appli
100	356.5	12.7	461	2	US-08-646-248-2	Sequence 2, Appli

101	356.5	12.7	461	5	PCT-US95-13924 2	Sequence 2, Appli
102	356.5	12.7	461	5	PCT-US95 13931 2	Sequence 2, Appli
103	351.5	12.5	764	3	US-09-144 759-20	Sequence 20, Appl
104	350.5	12.5	416	4	US-08 764-870-4	Sequence 4, Appli
105	350.5	12.5	416	4	US 08 980 115 4	Sequence 4, Appli
106	350.5	12.5	797	2	US-08-095-728B 2	Sequence 2, Appli
107	350.5	12.5	797	5	PCT-US92 02320A-2	Sequence 2, Appli
108	348.5	12.4	878	4	US-08-653-648A 15	Sequence 15, Appl
109	346.5	12.4	446	2	US-08-372 652 4	Sequence 3, Appli
110	346.5	12.4	446	5	PCT-US95-16311-3	Sequence 3, Appli
111	346	12.3	443	1	US-08 342-411A 4	Sequence 4, Appli
112	345.5	12.3	340	3	US 08 705-771-20	Sequence 20, Appl
113	344.5	12.3	446	4	US 08 776 844 2	Sequence 2, Appli
114	343.5	12.2	746	3	US-09-144 759 18	Sequence 18, Appl
115	335.5	12.0	460	1	US 08 342 411A 2	Sequence 2, Appli
116	325	11.6	323	2	US 09 132-619 4	Sequence 4, Appli
117	325	11.6	323	3	US-09 282-803B 4	Sequence 4, Appli
118	325	11.6	323	4	US 09 510-654 4	Sequence 4, Appli
119	325	11.6	557	4	US 08 653-648A 12	Sequence 12, Appl
120	324.5	11.6	575	4	US 08 653-648A 5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-139-617-1
; Sequence 1, Application US/09139617
; Patent No. 6222015
; GENERAL INFORMATION:
; APPLICANT: WILKINSON, HILARY
; TITLE OF INVENTION: ESTROGEN RECEPTOR
; FILE REFERENCE: 20047Y
; CURRENT APPLICATION NUMBER: US/09/139,617
; CURRENT FILING DATE: 1998-08-25
; EARLIER APPLICATION NUMBER: 60/058,271
; EARLIER FILING DATE: 1997-09-08
; EARLIER APPLICATION NUMBER: 60/060,520
; EARLIER FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 548
; TYPE: PRT
; ORGANISM: HUMAN
US-09-139-617-1

Query Match 100.0%; Score 2805; DB 4; Length 548;
Best Local Similarity 100.0%; Pred. No. 6.7e-292;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 78

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 138

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 198

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 258

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 318

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 378

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 438

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 498

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 548
```

RESULT 2
US-09-561-741A-1
; Sequence 1, Application US/09561741A
; Patent No. 6458551
; GENERAL INFORMATION:
; APPLICANT: WILKINSON, HILARY
; TITLE OF INVENTION: ESTROGEN RECEPTOR
; FILE REFERENCE: 20047Y
; CURRENT APPLICATION NUMBER: US/09/561,741A
; CURRENT FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 09/139,617
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/058,271
; PRIOR FILING DATE: 1997-09-08
; PRIOR APPLICATION NUMBER: 60/060,520
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 548
; TYPE: PRT
; ORGANISM: HUMAN
US-09-561-741A-1

Query Match 100.0%; Score 2805; DB 4; Length 548;
Best Local Similarity 100.0%; Pred. No. 6.7e-292;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 78

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 138

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 198

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 258

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 318

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 378

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 438

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 498

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 548
```

```
Db        439 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 498

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 548


RESULT 3
US-08-836-620A-3
; Sequence 3, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  485 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-836-620A-3

  Query Match             91.1%;  Score 2554;  DB 2;  Length 485;
  Best Local Similarity   99.8%;  Pred. No. 4.6e-265;
  Matches  484;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 120

Qy        166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCG 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCG 180

Qy        226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 240

Qy        286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 300

Qy        346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKIQHKEYLCVKAMIL 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKIQHKEYLCVKAMIL 360

Qy        406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 420

Qy        466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 480

Qy        526 NPQSQ 530
              |||||
Db        481 NLQSQ 485


RESULT 4
US-08-836-620A-2
; Sequence 2, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  485 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus rattus
US-08-836-620A-2

  Query Match             81.7%;  Score 2291;  DB 2;  Length 485;
  Best Local Similarity   88.7%;  Pred. No. 7.5e-237;
  Matches  430;  Conservative   23;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
              |||||||||||||:| : :||:||| | :|||||||| |||||  | | | ||||||||
Db          1 MTFYSPAVMNYSVPGSISNLDGGPVRLSISPNVLWPISGHLSPLAIHQQSSLLYAEPQKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVIGPGSKRDAHFCAVCSDYASGYHYGVWS 165
              ||||||||||||||||||||:|:| |||||  | :|||||| ||||||||||||||||
Db         61 PWCEARSLEHTLPVNRETLKRKLSGSSCASPVTSPNAKRDAHFCPVCSDYASGYHYGVWS 120

Qy        166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCG 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCG 180

Qy        226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVITLLEAEPPHVIISPP 285
              |||:|||||: |||||  |||||:|||||||:|||  |||||||||||||:||:||||
Db        181 YRIVRRQRSSSEQVHCISKAKRNGGHAPRVRELLLSILSPEQIVITLLEAEPPNVIVSRP 240
```

```
Qy          286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
                ||||||||||||||||||||||||| |||||||||||| |||||||||||||||:||||
Db          241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy          346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 360

Qy          406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
                |||||||| :| |:|:|||| |||||||||||||||||||||||:||||||||||||
Db          361 LNSSMYPLASANQEAESSRKLTHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHV 420

Qy          466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
                || |||||||||:|||||||||||||||||||| ||| |||:|||| ||||:|| ||
Db          421 RHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSECSSTEDSKNKESSQ 480

Qy          526 NPQSQ 530
                | |||
Db          481 NLQSQ 485


RESULT 5
US-08-836-620A-13
; Sequence 13, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607542.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  484 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus rattus
US-08-836-620A-13

  Query Match             81.5%;  Score 2286;  DB 2;  Length 484;
  Best Local Similarity   88.6%;  Pred. No. 2.6e-236;
  Matches  429;  Conservative   23;  Mismatches   32;  Indels    0;  Gaps    0;

Qy           46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
                |||||||||||:| : :||:||| | :|||||||| |||||  | | | ||||||||
Db            1 MTFYSPAVMNYSVPGSTSNLDGGPVRLSTSPNVIWPTSGHLSPLATHCQSSLLYAEPQKS 60

Qy          106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
                |||||||||||||||||||:|| |||||| | :|||||||| ||||||||||||||||
Db           61 PWCEARSLEHTLPVNRETLKRKLSGSSCASPVISPNAKRDAHFCPVCSDYASGYHYGVWS 120

Qy          166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCG 225
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCG 180

Qy          226 YRLVRRQRSADEQLHCAGKAKRSGCHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
                ||:|||||: ||:||  |||:|||||||:|||  |||||||||||||||:||:|||||
Db          181 YRIVRRQRSSSEQVHCLSKAKRNGGHAPRVKELLLSTLSPEQLVLTLLEAEPPNVLVSRP 240

Qy          286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
                | ||||||||||||||||||||| |||||||||||| |||||||||||||||:||||
Db          241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy          346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 360

Qy          406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
                |||||||| :| |:|:|||| |||||||||||||||||||||||:||||||||||||
Db          361 LNSSMYPLASANQEAESSRKLTHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHV 420

Qy          466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
                || |||||||||:|||||||||||||||||||| ||| |||:|||| ||||:|| ||
Db          421 RHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSECSSTEDSKNKESSQ 480

Qy          526 NPQS 529
                | ||
Db          481 NLQS 484


RESULT 6
US-08-836-620A-5
; Sequence 5, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607542.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  485 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-836-620A-5

  Query Match             80.8%;  Score 2267;  DB 2;  Length 485;
  Best Local Similarity   88.0%;  Pred. No. 2.8e-234;
  Matches  427;  Conservative   23;  Mismatches   35;  Indels    0;  Gaps    0;

Qy           46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
                | ||||||||||:|:  |||||| ||| |||||||| ||||||  | | | ||||||||
```

```
Db          1 MAFYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
              ||||||||||||||||||||||: 1: |||||| | :|||||||||||||||||||||||
Db         61 PWCEARSLEHTLPVNRETLKRKIGGSGCASPVTSPSTKRDAHFCAVCSDYASGYHYGVWS 120

Qy        166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACRLRKCYEVGMVKCGSRREKCG 225
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACRLRKCYEVGMVKCGSRREKCG 180

Qy        226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
              ||:||||||| ||:|| |||| || |||:||||::||||||||||||||||:|||:|||
Db        181 YRIVRRQRSASEQVHCLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRP 240

Qy        286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
              | |||||||||||||||||||||| |||||||||||| |||||||||||||||:||||
Db        241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy        346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMIL 360

Qy        406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
              |||||| | ||:|:|:|||| |||||||||||||:|| |||||||:||||||||||||
Db        361 LNSSMYHLATASQEAESSRKLTHLLNAVTDALVWVISKSRISSQQQSVRLANLLMLLSHV 420

Qy        466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
              || |||||||||:||||||||||||||||||| ||| |||:|| |   ||||||||||
Db        421 RHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSGCCSTEDSKSKEGSQ 480

Qy        526 NPQSQ 530
              | |||
Db        481 NLQSQ 485
```

```
RESULT 7
US-08-836-620A-14
; Sequence 14, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  484 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-836-620A-14
```

```
  Query Match             80.6%;  Score 2262;  DB 2;  Length 484;
  Best Local Similarity   88.0%;  Pred. No. 9.6e-234;
  Matches  426;  Conservative   23;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         46 MTFYSPAVMNYSVPSNVTNLEGGPGRQTISPNVLWPTPGRLSPLVVHRQLSHLYAEPQKS 105
              | |||||||||||:||:  |||||| ||| |||||||| |||||  | | | |||||||
Db          1 MAFYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
              ||||||||||||||||||||||: |: |||||| | :|||||||||||||||||||||||
Db         61 PWCEARSLEHTLPVNRETLKRKLGGSGCASPVTSPSTKRDAHFCAVCSDYASGYHYGVWS 120

Qy        166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCG 225
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACRLRKCYEVGMVKCGSRREKCG 180

Qy        226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
              ||:||||||| ||:|| |||| || |||:||||::||||||||||||||:|||:|||
Db        181 YRIVRRQRSASEQVHCLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRP 240

Qy        286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
              | |||||||||||||||||||||| |||||||||||| |||||||||||||||:||||
Db        241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy        346 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMIL 405
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        301 RSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTARFRELKLQHKEYLCVKAMIL 360

Qy        406 LNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
              |||||| | ||:|:|:|||| |||||||||||||:|| |||||||:||||||||||||
Db        361 LNSSMYHLATASQEAESSRKLTHLLNAVTDALVWVISKSRISSQQQSVRLANLLMLLSHV 420

Qy        466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQ 525
              || |||||||||:||||||||||||||||||| ||| |||:|| |   ||||||||||
Db        421 RHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGYKSSISGSGCCSTEDSKSKEGSQ 480

Qy        526 NPQS 529
              | ||
Db        481 NLQS 484
```

```
RESULT 8
US-08-836-620A-15
; Sequence 15, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  15:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  484 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
US-08-826-620A-15

  Query Match             71.9%;  Score 2017;  DB 2;  Length 484;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-207;
  Matches  483;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          86 LSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRD 145
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2 LSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRD 61

Qy         146 AHECAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR 205
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          62 AHECAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR 121

Qy         206 LRKCYEVGMVKCGSRREKCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSP 265
               |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         122 LRKCYEVGMVKCGSRREKCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSP 181

Qy         266 EQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLE 325
               |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         182 EQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLE 241

Qy         326 DQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTS 385
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         242 DQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTS 301

Qy         386 RFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSG 445
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         302 RFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSG 361

Qy         446 ISSQQQSMRLANLLMLLSHVRHA 468
               |||||||||||||||||||||||
Db         362 ISSQQQSMRLANLLMLLSHVRHA 384


RESULT 9
US-08-764-870-12
; Sequence 12, Application US/08764870
; Patent No. 6236946
;  GENERAL INFORMATION:
;     APPLICANT:  Scanlan, Thomas S
;     APPLICANT:  Baxter, John D
;     APPLICANT:  Fletterick, Robert J
;     APPLICANT:  Wagner, Richard L
;     APPLICANT:  Kushner, Peter J
;     APPLICANT:  Apriletti, James W
;     APPLICANT:  West, Brian
;     TITLE OF INVENTION:  Nuclear Receptor Ligands and Ligand
;     TITLE OF INVENTION:  Binding Domains
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Cooley Godward
;       STREET:  Five Palo Alto Square,  3000 El Camino Real
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94306
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/764,870
;       FILING DATE:  13-DEC-1996
;       CLASSIFICATION:  540
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/008,540
;       FILING DATE:  13-DEC-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/008,543
;       FILING DATE:  13-DEC-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/008,606
;       FILING DATE:  14-DEC-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nakamura, Jackie N
;       REGISTRATION NUMBER:  35,966
;       REFERENCE/DOCKET NUMBER:  UCAL-246/01US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (650)843-5000
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  595 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-764-870-12

  Query Match             44.0%;  Score 1233.5;  DB 4;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 2.9e-123;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy           5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
               |    || |     |  :||   | :|: ||   | :: | |  | | : | |  :
Db          21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy          61 NVINLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                 | |  ||| :                 : || :|   |  ||| |  | |    | |
Db          75 Q-IGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy         102 PQKSPWC-EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHECAVCSD 154
                 : | :   ||       |    :  |: |    :      |     | ::  :||||:|
Db         134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy         155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
               |||||||||||||:|||||||||||||:|||||||||||||||||||||||||||||||
Db         191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy         215 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR---------- - ---EL 258
               :| | |::| | |::: :|  |:      |: :   | |  :|
Db         251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy         259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
               |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||:|:::
Db         306 LALSLIADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy         317 PGFVELSLEDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
               ||||:|:| ||| |||  |:|:||:||:|||::|||||:|||:|:|||::|||||:||
Db         365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy         377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
               |||||||:|||| : || :|::|:|::|||||:|  :::| :  :    :  :|: :||
Db         425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy         436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                 |: ::||:|:: |||  ||| ||::|||:|| ||||||| :|||||||:|||||||
Db         485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy         496 NAHVL 500
               :||  |
Db         545 DAHRL 549


RESULT 10
US-08-980-115-12
; Sequence 12, Application US/08980115
; Patent No. 6266622
; GENERAL INFORMATION:
```

```
;  APPLICANT: Scanlan, Thomas S.
;  APPLICANT: Baxter, John D.
;  APPLICANT: Fletterick, Robert J.
;  APPLICANT: Wagner, Richard L.
;  APPLICANT: Kushner, Peter J.
;  APPLICANT: Apriletti, James W.
;  APPLICANT: West, Brian L.
;  APPLICANT: Shiau, Andrew K.
;  TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
;  FILE REFERENCE: UCAL-246/02US
;  CURRENT APPLICATION NUMBER: US/08/980,115
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: 08/764,870
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 60/008,606
;  EARLIER FILING DATE: 1995-12-14
;  EARLIER APPLICATION NUMBER: 60/008,543
;  EARLIER FILING DATE: 1995-12-13
;  EARLIER APPLICATION NUMBER: 60/008,540
;  EARLIER FILING DATE: 1995-12-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;  LENGTH: 595
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: DOMAIN
;  LOCATION: (287)..(549)
;  OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-12

  Query Match             44.0%;  Score 1233.5;  DB 4;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 2.9e-123;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EAPSLEHTLPVNRETLKREVSGNPCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||      |    :  |: |   :       |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRERGGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVP----------------EL 258
              :| | |::| | |::: :|  |:    |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGPGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAFPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  |:|||| :| |   |: ||:|||||  || |||:|||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:|||:|:|||:|||:|:|||::|||||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLIFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|::|:|||||  :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: :|||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
```

```
              :|| |
Db        545 DAHRL 549


RESULT 11
US-09-041-886-35
; Sequence 35, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Brodesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P 1JJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  595 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-041-886-35

  Query Match             43.7%;  Score 1224.5;  DB 4;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 2.7e-122;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EAPSLEHTLPVNRETLKREVSGNPCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||      |    :  |: |   :       |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRERGGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVP----------------EL 258
              :| | |::| | |::: :|  |:    |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGPGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAFPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  |:|||| :| |   |: ||:|||||  || |||:|||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364
```

```
Qy          417 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 476
                ||||:|:| ||| ||| |:|:|||:|||::|| ||:||||:|||::|||||||::||
Db          365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy          477 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLABLLNAVTD 535
                |||||||:|||| : || :|:::|:|::|||| :|  :::| :  :    :  :|: :||
Db          425 FDMLLATSSRFKMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy          536 ALVWVIAKSGTSSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 595
                 |: ::||:|:: ||| ||| ||:::|||:|| |||||||| :|||||||:|||||||
Db          485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy          596 NAHVL 600
                :|| |
Db          545 DAHRL 549


RESULT 12
US-08-453-998-2
; Sequence 2, Application US/08453998
; Patent No. 6444438
;  GENERAL INFORMATION:
;     APPLICANT:  CHAMBON, PIERRE
;     APPLICANT:  METZGER, DANIEL
;     APPLICANT:  WHITE, JOHN
;     TITLE OF INVENTION:  METHOD FOR THE PREPARATION OF A PROTEIN
;     TITLE OF INVENTION:  BY YEASTS USING AN INDUCIBLE SYSTEM, VECTORS AND
;     TITLE OF INVENTION:  CORRESPONDING TRANSFORMED STRAINS
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;       STREET:  1100 NEW YORK AVE., N.W.
;       CITY:  WASHINGTON
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20005
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/453,998
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/161,064
;       FILING DATE:  03 DEC-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CHAPIN, MARLANA K.
;       REGISTRATION NUMBER:  35,843
;       REFERENCE/DOCKET NUMBER:  1037/98493
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-861-3711
;       TELEFAX:  202-822-0944
;       TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  595 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-453-998-2

  Query Match             43.7%;  Score 1224.5;  DB 4;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 2.7e-122;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy            5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
                |    || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db           21 NELEPLNRP---   QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy           61 NVINLEGGPGRQ--   --  --------    TISPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                  | |  ||| :                 : || :|   |  ||| |  | |    | |
Db           75 QTIGLPYGPGSEAAAAPGSNGLGGEPPLNSVSPSPLMLLHPPPQLSPELQPHGQQVPYYLE 134

Qy          102 PQKSPWC--FAPSLFHTLPVNPETLKRKVSGNFCASPVTGPG   ---SKRDAHFCAVCSD 154
                : | :   ||       |    :  |:  |   :      |      | ::  :|||:|
Db          135 NEPSGYLVREAGPPAFYRP---NSDNRRQGGREELASLNDRGSMAMESAKETRYCAVCND 190

Qy          155 YASCYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDENRRKSCQACRLRKCYEVGM 214
                ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db          191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy          215 VKCGSKRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPKVR--   ---   ---   EL 258
                :| | |::| | |::: :|  |:     |: :   | |  :|      -
Db          251 MKGGIRKDRRGGRMLKHKRQRDD ----GEGEGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy          259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
                |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db          306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy          317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
                ||||:|:| ||| ||| |:|:|||:|||::|| ||:||||:|||::|||||||::||
Db          365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy          377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLABLLNAVTD 435
                |||||||:|||| : || :|:::|:|::|||| :|  :::| :  :    :  :|: :||
Db          425 FDMLLATSSRFKMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy          436 ALVWVIAKSGTSSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                 |: ::||:|:: ||| ||| ||:::|||:|| |||||||| :|||||||:|||||||
Db          485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy          496 NAHVL 500
                :|| |
Db          545 DAHRL 549


RESULT 13
US-08-836-620A-17
; Sequence 17, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  Orphan receptor
;     NUMBER OF SEQUENCES:  19
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/836,620A
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/EP96/03933
;       FILING DATE:
;       APPLICATION NUMBER:  GB 9518272.1
;       FILING DATE:  08-SEP-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9605550.4
;       FILING DATE:  15-MAR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9607582.0
;       FILING DATE:  11-APR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9609576.5
;       FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  591 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
```

```
;      ORGANISM:  Homo sapiens
US-08-836-620A-17

  Query Match             43.6%;  Score 1222.5;  DB 2;  Length 591;
  Best Local Similarity   48.1%;  Pred. No. 4.4e-122;
  Matches  262;  Conservative   89;  Mismatches  129;  Indels   65,  Gaps   14.

Qy           5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSFAVMNYSIPS 60
               |    || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db          21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy          61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                 | |  ||| :                 : || ::   |  ||| |  | |    | |
Db          75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPIMILHPPPQLSPFLQPHGQQVPYYLE 133

Qy         102 PQKSPWC--PAESLPHTLPVNFETIKRKVSGNPCASPVTGPG-----SKPDAHFCAVCSD 154
                : | :   ||       |    :  |:  |    :      |     | ::  :|||:|
Db         134 NRPSGYTVREAGPPAFYRP---NSDNRPQGGRLPLASTNDKGSMAMESAKETRYCAVCND 190

Qy         155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
               ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db         191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy         215 VKGGSRRFRGGYRLVPPQPSADEQLHCAGKAKRSGGHAPEVR----------------EL 258
               :| | |::| | |::: :|  |:     |: :   | |  :|
Db         251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy         259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
               |  :|: :|:|  |||:||| :| |   |: ||:|||||  || |||:|||||:|||::
Db         306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy         317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
               ||||:|:| ||| |||  |:|:||:||:|||::|| ||:|||:|:|||::|||||:|:|
Db         365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy         377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
               |||||||:|||| : || :|:::|:|::|||||  :|  :::| :  :  :  :|: :||
Db         425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy         436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
               |: ::||:|:: |||  ||| ||:::|||:|| |||||||| :|||||||:|||||||
Db         485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy         496 NAHVL 500
               :|| |
Db         545 DAHRL 549


RESULT 14
US-08-836-620A-16
; Sequence 16, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  596 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus rattus
US-08-836-620A-16

  Query Match             43.3%;  Score 1214.5;  DB 2;  Length 596;
  Best Local Similarity   45.5%;  Pred. No. 3.2e-121;
  Matches  261;  Conservative   100;  Mismatches  135;  Indels   77;  Gaps   17;

Qy           5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSSYVDSHHEYPAMIFY------SPAVMNY 56
               |    || |      |  :|:|   | :|: :|   :   ||    |      : |
Db          21 NELEPLNRP------QLKMPMERALGEVYVDNS-KPAVFNYPEGAAYEFNAAAAAAAAGA 73

Qy          57 SIP----SNVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSPLV---V 91
               | |    |::|    |||  :                 : || ::   | |:|| :
Db          74 SAPVYGQSSITY---GPGSEAAAFGANSLGAFPQLNSVSPSPIMILHPPPHVSPFLHPHG 130

Qy          92 HRQLSHLYAEPQKSPWCEARSLEHTLP---VNRETLKRKVSGNRCASPVIGPG-----SK 143
               |:   :|  ||       | ::  | |       :  |: :|    |  : |     |
Db         131 HQVPYYLENEP------SAYAVRDTGPPAFYRSNSDNRPQNGRERLSSSSEKGNMIMESA 184

Qy         144 PDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQA 203
               ::  :|||:|||||||||||||||||||||||||||||||||:|||||||||||||||||
Db         185 KETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCIIDKNRRKSCQA 244

Qy         204 CRLRKCYEVGMVKGGSRRFRGGYRLVPPQPSADE-----QIHCAGKAKRSG--------G 250
               ||||||||||:| | |::| | |::: :|  |:     ::  :|  : :
Db         245 CRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDLEGPNEMGTSGDMRAANLWPSPLVIK 304

Qy         251 HAPRVRELLLDALSPEQLVLTLLEAEPPHVLIS-RPSAPFTEASMMMSLTKLADKELVHM 309
               |  :   |  :|: :|:|  ||:|||| :     || ||:|||||  || |||:|||||
Db         305 HTKKNSPAL--SLTADQMVSALLDAEPPLIYSEYDPSKPFSEASMMGLLTNLADRELVHM 362

Qy         310 ISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKC 369
               |:|||:|||  :|:|  ||| |||  |:|:||:||:|||::|| ||:|||:|:|||::|
Db         363 TNWAKRVPGFGDLNLHDQVHLLECAWLEILMIGLVWRSMEHPGKLIFAPNLLLDRNQGKC 422

Qy         370 VEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAH 428
               |||::||||||||:|||| : || :|:::|:|::|||||  :|  :::| :  :  :
Db         423 VEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYIFLSSTLKSLEEKDHIHR 482

Qy         429 LLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVY 488
               :|: : | |: ::||:|::  |||  ||| ||::|||:|| ||||||| ||||||||:|
Db         483 VLDKINDTIIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLY 542

Qy         489 DLLLEMLNAHVLPGCKSSTIGSEGSPAEDSKSK 521
               |||||||:|| |    | :     | | |:|:
Db         543 DLLLEMLDAHRLHAPASRM---GVPPEEDSQSQ 572


RESULT 15
5223606-5
;Patent No. 5223606
;    APPLICANT:  BLAUDIN DE THE, HUGHES;MARCHIO, AGNES;TIOLLAIS,
;PIERRE;DEJEAN, ANNE
;    TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED
;PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA
;    NUMBER OF SEQUENCES: 11
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/134,130
;      FILING DATE: 17-DEC-1987
;    PRIOR APPLICATION DATA:
;SEQ ID NO:5:
;      LENGTH: 410
```

5223606-5

Query Match 41.4%; Score 1160; DB 6; Length 410;
Best Local Similarity 55.0%; Pred. No. 1.2e-115;
Matches 219; Conservative 70; Mismatches 56; Indels 26; Gaps 5;

```
Qy      149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
            ||||:|||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db        1 CAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRK 60

Qy      209 CYEVGMVKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR------- 256
            |||||||:| | |:|: | |::: :|  |:     |: :  | |  :|
Db       61 CYEVGMMKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSACDMRAANLWPSPLMIK 115

Qy      257 -   ELLLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMI 310
                |  :|: :|:|  ||:|||  :| |   |: ||:|||||  || |||:||||||
Db      116 RSKKNSLALSLTADQMVSALLDAEP- ILYSEYDPTRPFSEASMMGLLTNLADRELVHMI 173

Qy      311 SWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCV 370
            :|||::||||:|:| ||| |||  |:|:||:||:|||::|| ||:||:|:|||::|||
Db      174 NWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCV 233

Qy      371 EGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHL 429
            ||:|||||||||:|||| : || :|:::|:|::|||| :|  :::| :  :    :  :
Db      234 EGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRV 293

Qy      430 LNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYD 489
            |: :|| |: ::||:|:: |||  ||| ||::|||:|| |||||||| :||||||:||
Db      294 LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYD 353

Qy      490 LLLEMLNAHVL 500
            |||||||:|| |
Db      354 LLLEMLDAHRL 364
```

RESULT 16
US-09-249-645-1
; Sequence 1, Application US/09249645
; Patent No. 6228990
; GENERAL INFORMATION:
; APPLICANT: Ljunqqren, Jan
; APPLICANT: Thorsell, Ann-Gerd
; APPLICANT: Enqstrom, Owe
; APPLICANT: Bonn, Tomas
; APPLICANT: Carlquist, Mats
; APPLICANT: Brzozowski, Andrzeji
; APPLICANT: Pike, Ashley
; APPLICANT: Hubbard, Roderick
; TITLE OF INVENTION: ESTROGEN RECEPTOR LIGANDS
; FILE REFERENCE: 101987-100
; CURRENT APPLICATION NUMBER: US/09/249,645
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: GB 9803062.0
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Human
US-09-249-645-1

Query Match 40.2%; Score 1126.5; DB 4; Length 229;
Best Local Similarity 95.0%; Pred. No. 1.8e-112;
Matches 228; Conservative 0; Mismatches 1; Indels 11; Gaps 2;

```
Qy      261 DALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
            ||||||||||||||||||||||||||      |||||||||||||||||||||||||||
Db        1 DALSPEQLVLTLLEAEPPHVLISRPS-   --ASMMMSLTKLADKELVHMISWAKKIPGFV 55

Qy      321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       56 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 115

Qy      381 LATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWV 440
            ||||||||||||||||||||||||||||||||||||      |||||||||||||||||||
Db      116 LATTSRFRELKLQHKEYLCVKAMILLNSSMYPLV   - - DSSRKLAHLLNAVTDALVWV 169

Qy      441 IAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
            ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      170 IAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLMMKCKNVVPVYDLLLEMLNAHVL 229
```

RESULT 17
US-09-249-645-2
; Sequence 2, Application US/09249645
; Patent No. 6228990
; GENERAL INFORMATION:
; APPLICANT: Ljunqqren, Jan
; APPLICANT: Thorsell, Ann-Gerd
; APPLICANT: Enqstrom, Owe
; APPLICANT: Bonn, Tomas
; APPLICANT: Carlquist, Mats
; APPLICANT: Brzozowski, Andrzeji
; APPLICANT: Pike, Ashley
; APPLICANT: Hubbard, Roderick
; TITLE OF INVENTION: ESTROGEN RECEPTOR LIGANDS
; FILE REFERENCE: 101987-100
; CURRENT APPLICATION NUMBER: US/09/249,645
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: GB 9803062.0
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Rat
US-09-249-645-2

Query Match 37.9%; Score 1064; DB 4; Length 228;
Best Local Similarity 90.2%; Pred. No. 9e-106;
Matches 212; Conservative 8; Mismatches 7; Indels 8; Gaps 1;

```
Qy      263 LSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVEL 322
            ||||||||||||||||:||:|||| ||||||||||||||||||||| ||||||||||
Db        2 LSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVEL 61

Qy      323 SLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA 382
            || ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       62 SLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA 121

Qy      383 TTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIA 442
            |||||||||||||||||||||||||||||| :| |:|:|||| ||||||||||||||
Db      122 TTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLTHLLNAVTDALVWVIA 181

Qy      443 KSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
            |||||||||:|||||||||| |||||||||:||        |||||||||
Db      182 KSGISSQQQSVRLANLLMLLSHVRHTSNKGMEHLLSMK-- -- --DLLLEMLNA 228
```

RESULT 18
US-08-836-620A-7
; Sequence 7, Application US/08836620A
; Patent No. 5958710
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Orphan receptor
; NUMBER OF SEQUENCES: 19
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,620A

```
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/EP96/03933
;       FILING DATE:
;       APPLICATION NUMBER:  GB 9518272.1
;       FILING DATE:  08-SEP-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9605550.4
;       FILING DATE:  15-MAR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9607532.0
;       FILING DATE:  11-APR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9609576.5
;       FILING DATE:  08-MAY-1996
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  226 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  Rattus rattus
US-08-836-620A-7

  Query Match             37.6%;  Score 1055;  DB 2;  Length 226;
  Best Local Similarity   91.2%;  Pred. No. 8.2e-105;
  Matches  206;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        305 ELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDR 364
              ||||||| |||||||||||||| |||||||||||||||:|||||||||||||||||||||
Db          1 ELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDR 60

Qy        365 DEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSR 424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| :| |:|:|||
Db         61 DEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSR 120

Qy        425 KLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNV 484
              || |||||||||||||||||||||||||:|||||||||||||| ||||||||||:|||||||
Db        121 KLTHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNV 180

Qy        485 VPVYDLLLEMLNAHVLRGCKSSITGSESPAEDSKSKEGSQNPQSQ 530
              |||||||||||||| ||| |||||:|||||  ||||:|| ||| |||
Db        181 VPVYDLLLEMLNAHTLRGYESSISGSESSSTEDSKNKESSQNLQSQ 226


RESULT 19
US-08-564-264-1
; Sequence 1, Application US/08564264
; Patent No. 6040430
;  GENERAL INFORMATION:
;     APPLICANT:  STEWART, Francis
;     TITLE OF INVENTION:  REGULATION OF SITE-SPECIFIC
;     TITLE OF INVENTION:  RECOMBINATION BY SITE-SPECIFIC RECOMBINASE/NUCLEAR
;     TITLE OF INVENTION:  RECEPTOR FUSION PROTEINS
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram
;       STREET:  655 Fifteenth Street N.W. Suite 330
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005-5701
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/564,264
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/EP94/02088
;       FILING DATE:  28-JUN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  EP 93 110 298.2
;       FILING DATE:  28-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Murray, Robert B.
;       REGISTRATION NUMBER:  22,980
;       REFERENCE/DOCKET NUMBER:  P564-5019
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)638-5000
;       TELEFAX:  (202)638-4810
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  773 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (1-423)
;       OTHER INFORMATION:  /note= "FLP recombinase domain."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (424-428)
;       OTHER INFORMATION:  /note= "Linker peptide."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (429-773)
;       OTHER INFORMATION:  /note= "Estrogen binding domain."
US-08-564-264-1

  Query Match             28.9%;  Score 810;  DB 3;  Length 773;
  Best Local Similarity   36.9%;  Pred. No. 1.2e-77;
  Matches  198;  Conservative   91;  Mismatches  120;  Indels  128;  Gaps   16;

Qy          5 NSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVTN 64
              |   | |      |   ||  :::|    | |::  |     || :    ::    : :||
Db        278 NLVRSYNKALKKNAPYSIFAIKNG----PKSHIGRH----LMTSF----LSMKGLTELTN 325

Qy         65 LEG-------GPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEAKSLEHTL 117
              : |           :||  : :   | |   ||      |  ||
Db        326 VVGNWSDKRASAVARTTYTHQITAIPDHYFALV-----SRYYAYD-------------- 365

Qy        118 PVNRFTIKPKVSGNPCASPV----------TGPGSKRDAHFCAVCS----DYASGYHYG 162
              |:::| :  |   |    |:           :  || |   :  : |    || | |
Db        366 PISKFMIALKDETN----PIFEWQHIEQLKGSAFGSIRYPAWNGIISQFVLDYISSY--- 418

Qy        163 VWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRE 222
                                                ||| | :          | :| | |::
Db        419 --------------------------------INRRISVR--------  --GSMKGCIRKD 436

Qy        223 PCCYRLVPPQRSADEQLHCACKAFRSGCHAPPVP------- --- ----FTIIDAISPE 266
              | | |::: :|  |:     |: :   | )  :|                    |  :|: :
Db        437 PRCQRMLEHFPPIGV-----CPSRCEVCSACTDMRAANIWPSPLMIKRSKKNSIALSLTAD 491

Qy        267 QLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSL 324
              |:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||::||||:|:|
Db        492 QMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTL 550

Qy        325 FDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATT 384
              ||| ||| |:|:||:||:|||::|| ||:|||:|:|||::|||||||::|||||||:
Db        551 HDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATS 610

Qy        385 SRFRELKLQHKEYLCVKAMILLNsSMYPLVIAT-QDADSSPKLAHLLNAVTDALVWVIAK 443
              |||| : || :|:|:|:|:|||| :|  :::|  :  :     :  :|: :|| |: :||
Db        611 SPFPMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAK 670

Qy        444 SGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
              :|:: ||| |:| ||::|||:|| |||||||| :|||||||:|||||||:|| |
Db        671 AGITQQQHQPLAQFILILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRL 727


RESULT 20
```

```
US 08 693 940 3
; Sequence 3, Application US/08693940
; Patent No. 6133027
;  GENERAL INFORMATION:
;    APPLICANT:  Yee, Jiing Kuan
;    APPLICANT:  Friedman, Theodore
;    APPLICANT:  Chen, Shin Tai
;    TITLE OF INVENTION:  Inducible Expression System
;    TITLE OF INVENTION:  Useful in the Generation of Packaging Cell Lines for
;    TITLE OF INVENTION:  Pseudotyped Retroviral Vectors
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/693,940
;      FILING DATE:  07-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Francis, Carol L.
;      REGISTRATION NUMBER:  34,513
;      REFERENCE/DOCKET NUMBER:  6510-055001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650 327 3400
;      TELEFAX:  650 327 3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  651 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US 08 693 940 3

  Query Match             27.2%;  Score 762;  DB 4;  Length 651;
  Best Local Similarity   59.1%;  Pred. No. 1.3e-72;
  Matches  143;  Conservative   57;  Mismatches   38;  Indels    4;  Gaps    3;

Qy        262 ALSPEQLVLTLLEAEPPHVLISR-PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGF 319
              :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||::|||
Db        465 SLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGF 523

Qy        320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
              |:|:| ||| |||  |:|:||:||:|||::|| ||:|||:|:|||::|||||::|||||
Db        524 VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDM 583

Qy        380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALV 438
              ||||:|||| : || :|::|:|::|||||| :|  :::| :  :     :  :|: :|| |:
Db        584 LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLI 643

Qy        439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
              ::|:|:: |||  ||| |||:|||:|| |||||||| :|||||||:|||||||:||
Db        644 HLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAH 703

Qy        499 VL 500
               |
Db        704 RL 705
```

```
RESULT 21
US 09 566 660 3
; Sequence 3, Application US/09566660
; Patent No. 6432705
; GENERAL INFORMATION:
;  APPLICANT:  Yee, Jiing-Kuan
;  APPLICANT:  Friedmann, Theodore
;  APPLICANT:  Chen, Shin-Tai
;  TITLE OF INVENTION:  Inducible Expression System
;  FILE REFERENCE:  6510-055CON
;  CURRENT APPLICATION NUMBER: US/09/566,660
;  CURRENT FILING DATE:  2000-05-08
;  PRIOR APPLICATION NUMBER:  08/693,940
;  PRIOR FILING DATE:  1996-08-07
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 651
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Multi-chimeric transactivating factor
US-09-566-660-3

  Query Match             27.2%;  Score 762;  DB 4;  Length 651;
  Best Local Similarity   59.1%;  Pred. No. 1.3e-72;
  Matches  143;  Conservative   57;  Mismatches   38;  Indels    4;  Gaps    3;

Qy        262 ALSPEQLVLTLLEAEPPHVLISR-PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGF 319
              :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||::|||
Db        465 SLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGF 523

Qy        320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
              |:|:| ||| |||  |:|:||:||:|||::|| ||:|||:|:|||::|||||::|||||
Db        524 VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDM 583

Qy        380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT QDADSSRKLAHLLNAVTDALV 438
              ||||:|||| : || :|::|:|::|||||| :|  :::| :  :     :  :|: :|| |:
Db        584 LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLI 643

Qy        439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
              ::|:|:: |||  ||| |||:|||:|| |||||||| :|||||||:|||||||:||
Db        644 HLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAH 703

Qy        499 VL 500
               |
Db        704 RL 705
```

```
RESULT 22
US-08-826-620A-9
; Sequence 9, Application US/08826620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/826,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-836-620A-9

  Query Match             23.8%;  Score 668.5;  DB 2;  Length 243;
  Best Local Similarity   54.9%;  Pred. No. 2.7e-63;
  Matches  124;  Conservative   52;  Mismatches   49;  Indels    1;  Gaps    1;

Qy        305 ELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDR 364
              |||||:|||::||| :|:| ||| ||| |:|:||:||:||::|||||:|||:|:|||
Db          1 ELVHMINWAKRVPGFGDLNLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDR 60

Qy        365 DEGKCVEGILEIFDMLLATTSRFRFLKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSS 423
              ::|||||::|||||||||:|||| : || :|:|:|:|::|||| :|  :::| :  :
Db         61 NQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEK 120

Qy        424 RKLAHLLNAVTDAIVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKN 483
                :  :|: : | |: ::||:|:: ||| |||:|::|||:|| ||||||| ||||||
Db        121 DHIHRVLDKITDTLIHLMAKAGLTLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKN 180

Qy        484 VVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSK 521
              |||:|||||||:|| |    | :      | | |:|:
Db        181 VVPLYDLLLEMLDAHRLHAPASRM---GVPPEEPSQSQ 215


RESULT 24
US-08-836-620A-10
; Sequence 10, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-836-620A-10

  Query Match             23.5%;  Score 660.5;  DB 2;  Length 243;
  Best Local Similarity   60.4%;  Pred. No. 1.9e-62;
  Matches  119;  Conservative   48;  Mismatches   29;  Indels    1;  Gaps    1;

Qy        305 ELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDR 364
              |||||:|||::|||| :|:| ||| ||| |:|:||:||:||::|| ||:|||:|:|||
Db          1 ELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDR 60

Qy        365 DEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSS 423
              ::|||||::|||||||||:|||| : || :|:|:|:|::|||| :|  :::| :  :
Db         61 NQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEK 120
```

```
Qy         424 RKLAHLTNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKN 483
               :  :|: :|| | : ::|||:|:: ||| |||| ||::|||:|| ||||||||| :|||||
Db         121 DHIHRVLDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKN 180

Qy         484 VVPVYDLLLEMLNAHVL 500
               |||:|||||||||:|| |
Db         181 VVPLYDLLLEMLDAHRL 197


RESULT 25
US-09-040-508-2
; Sequence 2, Application US/09040508
; Patent No. 6069239
;  GENERAL INFORMATION:
;    APPLICANT:  Mathias, Stephen L.
;    TITLE OF INVENTION:  NOVEL COMPOUNDS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,508
;      FILING DATE:  17-MAR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9705451.4
;      FILING DATE:  17-MAR-1997
;      APPLICATION NUMBER:  GB 9803289.9
;      FILING DATE:  16-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-80008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  435 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-040-508-2

  Query Match             23.1%;  Score 648;  DB 4;  Length 435;
  Best Local Similarity   35.1%;  Pred. No. 1.1e-60;
  Matches  174;  Conservative   93;  Mismatches  145;  Indels   84;  Gaps   23;

Qy          17 NCSQSTLPLEHGSTYTPSSYVDS-HHEYPAMTEYSPAVMNYSTPSNVTNLEGGPGRQTTS 75
               :|| | :  |  |   |:|  || :|  |  :  | |  :|| |  :   : |
Db          11 SCS-SFIKTEPSS---PASLTDSVNHHSPGGS--SDASGSYS--STMNGHQNGLDSPPLY 62

Qy          76 PNVLWPTPGHLSPLVVHRQL-----SHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSG 130
               |:   |  |   |:   |:|     | :  :||    ||   : :::|
Db          63 PSA--PTLGGSGPV---RKLYDDCSSTIVEDPQTK--CE--YMLNSMP---------- 101

Qy         131 NRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
                        ||     | || | |||||||| ||| |||||||:||| :| |||||:
Db         102 ------------KR---LCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNE 146

Qy         191 CTIDKNRRKSCQACRLRKCYEVGMVKGSR--RERCGYRLVRRQRSADEQLHCACKAKRS 248
               | | | |||||||||  || :|||:| | |  | | | :  :|:  |:
Db         147 CEIIKRERKESCQACRFMKCLKVGMLKEGVRLDRVRGGRQKYKRRIDAE---------- 194

Qy         249 GGHAPRVR-ELLLDALSP-EQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLTKLADKE 305
               ::| :  :|:  |  | ::|  || |||  :  :  |: | ::   : :|  |||:|
Db         195 -NSPYLNPQLVQPAKKPYNKIVSHLLVAEPEKIYAMPDPTVPDSDIKALTTLCDLADRE 252

Qy         306 LVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRD 365
               || :| ||| ||||  ||| ||: ||:| |||:|:::|::||:    :|:|| | ::| |
Db         253 LVVIIGWAKHIPGFSTLSLADQMSLLQSAWMEILIIGVVYRSLSFEDELVYADDYIMDED 312

Qy         366 EGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSR 424
               : | : |:|:: : :|    ::: :||: :|:: :||: | || ||:       :| :: :
Db         313 QSK-LAGLLDLNNAILQLVKKYKSMKLEKEEFVTLKAIALANSDSMH-  -  IEDVEAVQ 366

Qy         425 KLAHLLN-AVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKN 483
               ||  :|: |: |       ::|    :   |   :|| |  :|  | | ::|  |:| :
Db         367 KLQDVLHEALQD-  ---YEAG-QHMEDPKRAGKMLMTLPLLRQTSTKAVQHFYNTKLEG 419

Qy         484 VVPVYDLLLEMLNAHV 499
                |::: | |||| | |
Db         420 KVPMHKLFLEMLEAKV 435


RESULT 26
US-09-500-654-2
; Sequence 2, Application US/09500654
; Patent No. 6359116
; GENERAL INFORMATION:
;  APPLICANT: Stephen L. Mathias
;  TITLE OF INVENTION: No. 6359116el Compounds
;  FILE REFERENCE: GP-30008-D1
;  CURRENT APPLICATION NUMBER: US/09/500,654
;  CURRENT FILING DATE: 2000-02-09
;  EARLIER APPLICATION NUMBER: 9705451.4
;  EARLIER FILING DATE: 1997-03-17
;  EARLIER APPLICATION NUMBER: 9803289.9
;  EARLIER FILING DATE: 1998-02-16
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 435
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-500-654-2

  Query Match             23.1%;  Score 648;  DB 4;  Length 435;
  Best Local Similarity   35.1%;  Pred. No. 1.1e-60;
  Matches  174;  Conservative   93;  Mismatches  145;  Indels   84;  Gaps   23;

Qy          17 NCSQSTLPLEHGSTYTPSSYVDS-HHEYPAMTEYSPAVMNYSTPSNVTNLEGGPGRQTTS 75
               :|| | :  |  |   |:|  || :|  |  :  | |  :|| |  :   : |
Db          11 SCS-SFIKTEPSS---PASLTDSVNHHSPGGS--SDASGSYS--STMNGHQNGLDSPPLY 62

Qy          76 PNVLWPTPGHLSPLVVHRQL-----SHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSG 130
               |:   |  |   |:   |:|     | :  :||    ||   : :::|
Db          63 PSA--PTLGGSGPV---RKLYDDCSSTIVEDPQTK--CE--YMLNSMP---------- 101

Qy         131 NRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
                        ||     | || | |||||||| ||| |||||||:||| :| |||||:
Db         102 ------------KR---LCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNE 146

Qy         191 CTIDKNRRKSCQACRLRKCYEVGMVKGSR--RERCGYRLVRRQRSADEQLHCACKAKRS 248
               | | | |||||||||  || :|||:| | |  | | | :  :|:  |:
Db         147 CEIIKRERKESCQACRFMKCLKVGMLKEGVRLDRVRGGRQKYKRRIDAE---------- 194

Qy         249 GGHAPRVR-ELLLDALSP-EQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLTKLADKE 305
               ::| :  :|:  |  | ::|  || |||  :  :  |: | ::   : :|  |||:|
Db         195 -NSPYLNPQLVQPAKKPYNKIVSHLLVAEPEKIYAMPDPTVPDSDIKALTTLCDLADRE 252

Qy         306 LVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRD 365
               || :| ||| ||||  ||| ||: ||:| |||:|:::|::||:    :|:|| | ::| |
Db         253 LVVIIGWAKHIPGFSTLSLADQMSLLQSAWMEILIIGVVYRSLSFEDELVYADDYIMDED 312

Qy         366 EGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSR 424
               : | : |:|:: : :|    ::: :||: :|:: :||: | || ||:       :| :: :
Db         313 QSK-LAGLLDLNNAILQLVKKYKSMKLEKEEFVTLKAIALANSDSMH-  -  IEDVEAVQ 366

Qy         425 KLAHLLN-AVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKN 483
               ||  :|: |: |       ::|    :   |   :|| |  :|  | | ::|  |:| :
Db         367 KLQDVLHEALQD-  ---YEAG-QHMEDPKRAGKMLMTLPLLRQTSTKAVQHFYNTKLEG 419

Qy         484 VVPVYDLLLEMLNAHV 499
                |::: | |||| | |
Db         420 KVPMHKLFLEMLEAKV 435
```

```
Qy        366 EGKCVFGILFIFDMLLATTSRFRELKLQHKEYLCVKAMILLNS-SMYPLVTATQDADSSR 424
              : | : |:||:: : :|    ::: :||: :|:: :||: | || ||:       :| :: :
Db        313 QSK-IAGILDLNNAILQLVKKYKSMKLEKEEFVTLKALALANSDSMH-----IEDVEAVQ 366

Qy        425 KLAHLLN-AVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKN 483
              || :|: |: |       ::|    :   |   :|| |  :|  | | ::|  |:| :
Db        367 KLQDVLHEALQD------YEAG-QHMFDPPRAGKMLMTLPLLRQTSTKAVQHFYNIKLEG 419

Qy        484 VVPVYDLLLEMLNAHV 499
               |||: | |||| | |
Db        420 KVPMHKLFLEMLEAKV 435


RESULT 27
US-09-141-000-4
; Sequence 4, Application US/09141000
; Patent No. 6054295
; GENERAL INFORMATION:
;  APPLICANT: Chen, Fang
;  TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR
;  TITLE OF INVENTION: RECEPTOR PROTEINS
;  FILE REFERENCE: 19999Y
;  CURRENT APPLICATION NUMBER: US/09/141,000
;  CURRENT FILING DATE: 1998-08-26
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: Human
US-09-141-000-4

  Query Match             23.1%;  Score 648;  DB 3;  Length 458;
  Best Local Similarity   35.1%;  Pred. No. 1.2e-60;
  Matches  174;  Conservative   93;  Mismatches  145;  Indels   84;  Gaps   23;

Qy         17 NCSQSILPLEHGSIYIPSSYVDS-HHEYPAMTFYSPAVMNYSIPSNVTNLEGGPSRQTTS 75
              :|| : :  |  |   |:|  || :|  | :  : | |  :|| | :   : |
Db         34 SCS-SFIKTEPSS---PASLIDSVNHHSPGGS--SDASGSYS--STMNGHQNGLDSPPLY 85

Qy         76 PNVLWPTPGHLSPLVVHRQL-----SHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSG 130
              |:   |  |   |:   |:|     | :  :||   ||   : :::|
Db         86 PSA--PILGGSGPV---RKLYDDCSSTIVEDPQTK--CE--YMLNSMP------------ 124

Qy        131 NRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
                        ||    | || | |||||||| ||| ||||||||:|||: :| |||||:
Db        125 ------------KR---LCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNE 169

Qy        191 CTIDKNRRKSCQACRLRKCYEVGMVKCGSR--REPCGYRLVPRQRSADEQLHCACKAKRS 248
              | | | |||||||||  || :|||:| | |  | | | :  :|:  |:
Db        170 CPIEKRRRKSCQACPEMKCLKVGMLKEGVRLDPVPGCRQKYKRRPIDAF-            217

Qy        249 GGHAPRVR-FLILDAISP-PQLVITLFAEPPHV-LISRPSAPFTEASMMMSLTKLADKEL 305
              ::| :  :|:  |  | ::|  || |||  :  :  |: | ::    : :|  |||:|
Db        218 --NSPYLNPQLVQPAKKPYNKIVSHLLVAEPEKIYAMPDFTVPDSDIKALTTLCDLADRE 275

Qy        306 LVHMISWAKKIPGFVFLSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRD 365
              || :| ||| |||| ||| || :|:| ||||:::|:::||:    :|:::| | ::| |
Db        276 LVVIIGWAKHIPGFSTLSLADQMSLLQSAWMEILILGVVYRSLSFEDELVYADDYIMDED 335

Qy        366 EGKCVFGILFIFDMLLATTSRFRELKLQHKEYLCVKAMILLNS-SMYPLVTATQDADSSR 424
              : | : |:||:: : :|    ::: :||: :|:: :||: | || ||:       :| :: :
Db        336 QSK-IAGILDLNNAILQLVKKYKSMKLEKEEFVTLKALALANSDSMH-----IEDVEAVQ 389

Qy        425 KLAHLLN-AVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKN 483
              || :|: |: |       ::|    :   |   :|| |  :|  | | ::|  |:| :
Db        390 KLQDVLHEALQD------YEAG-QHMFDPPRAGKMLMTLPLLRQTSTKAVQHFYNIKLEG 442

Qy        484 VVPVYDLLLEMLNAHV 499
               |||: | |||| | |
Db        443 KVPMHKLFLEMLEAKV 458


RESULT 28
US-09-141-000-2
; Sequence 2, Application US/09141000
; Patent No. 6054295
; GENERAL INFORMATION:
;  APPLICANT: Chen, Fang
;  TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR
;  TITLE OF INVENTION: RECEPTOR PROTEINS
;  FILE REFERENCE: 19999Y
;  CURRENT APPLICATION NUMBER: US/09/141,000
;  CURRENT FILING DATE: 1998-08-26
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 500
;   TYPE: PRT
;   ORGANISM: Human
US-09-141-000-2

  Query Match             22.5%;  Score 630;  DB 3;  Length 500;
  Best Local Similarity   39.4%;  Pred. No. 1.2e-58;
  Matches  141;  Conservative   70;  Mismatches  108;  Indels   40;  Gaps    9;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | || | |||||||| ||| |||||||:|||: :| |||||:| | | |||||||| |
Db        103 CLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPAINECEIIKRERRKSCQACRFMK 162

Qy        209 CYEVGMVKCGSR--REPCGYRLVPRQRSADEQLHCACKAKRSGGHAPRVRFLILDAISPF 266
              | :|||:| | |  | | | :  :|:  ::  :                 | |    |
Db        163 CLKVGMLKEGVRLDRVRGGRQKYKRRLDSESSPY-------------- -ISLQISPPA 205

Qy        267 QLVLT-----LLEAEPPHV-LISPPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
              :  ||     || ||| .  . | |  :   : :|  |||:||| :| ||| ||||
Db        206 KEPLIKLVSYLLVAEPDELYAMPFPGMFEGDIKALITLCDLADRELVVIIGWAKHIPGFS 265

Qy        321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
              ||| ||. ||:| |||.|..|...||: .  || .| | ::| ·  · · |:||·· ·
Db        266 SLSLGDQMSLLQSAWMEILILGIVYRSLPYDDKLVYAEDYIMDEEHSR-LAGLLELYRAI 324

Qy        381 LATTSRFRELKLQHKEYLCVKAMILLNS-SMYPLVTATQDADSSRKLAHLLN-AVTDALV 438
              |    |:::|||: :|:: :||: | || |||       :| :: :||  ||: |: | :
Db        325 LQLVRRYKKLKVEKEEFVTLKALALANSDSMY-----IEDLEAVQKLQDLLHEAIQDYEL 379

Qy        439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
                        ::  |  ||: |  :|  : | ::|  ::| :  |||: | |||| |
Db        380 -------SQRHEEPWRTGKLLLTLPLLRQTAAKAVQHFYSVKLQGKVPMHKLFLEMLEA 431


RESULT 29
US-08-826-620A-19
; Sequence 19, Application US/08826620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/826,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9607532.0
;       FILING DATE:  11-APR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9609576.5
;       FILING DATE:  08-MAY-1996
;   INFORMATION FOR SEQ ID NO:  19:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  431 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
US-08-836-620A-19

  Query Match             22.4%;  Score 627.5;  DB 2;  Length 431;
  Best Local Similarity   39.3%;  Pred. No. 1.8e-58;
  Matches  141;  Conservative   69;  Mismatches  109;  Indels   40;  Gaps    9;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | || | ||||||||| ||| ||||||||:|||: :| ||||:| | | ||||||||| |
Db        103 CLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNECEITKRRRKSCQACRFMK 162

Qy        209 CYEVGMVKGGSR--RERGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLLDALSPE 266
              | :|||:| | | | | | :  :|:  ::   :                 | |    |
Db        163 CIKVGMLKEGVRLDRVRGGRQKYKRRLDSENSPY ---  ----------- LSLQISPPA 205

Qy        267 QLVLT    --LLEAEPPHVLISRP-SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
              :  ||      || |||  :    |   |  :    : :|  |||:||| :||||| ||||
Db        206 KKPLTKIVSYLLVAEPDKLYAMPPDEVPECDIKAITTLCDLADRELVFLISWAKHIPGFS 265

Qy        321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
              |:| ||: ||:| |||:|:::|:::|| :  || :| | ::|  :  : | |:||::  :
Db        266 NLTLGDQMSLLQSAWMEILILGIVYRSLPYDDKLAYAEDYIMDEEHSRLV-GLLELYRAI 324

Qy        381 LATTSRFRELKLQHKEYLCVKAMILLNS-SMYPLVTATQDADSSRKLAHLLN-AVTDALV 438
              |    |:::||:: :|:: :||: | || |||       :: :: :||  ||: |: |  :
Db        325 LQLVRKYKKLKVEKEEFVMLKAIALANSDSMY-----IENLEAVQRLQDLLHEAIQDYEL 379

Qy        439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
                          ::  |   || : |  :|  : | ::|  ::| :  |||: | |||| |
Db        380 --      SQRHEEPRRAGKLLLTLPLLRQTAAKAVQHFYSVKLQGKVPMHKLFLEMLEA 431


RESULT 30
US-08-836-620A-18
; Sequence 18, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  Orphan receptor
;     NUMBER OF SEQUENCES:  19
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/836,620A
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/EP96/03933
;       FILING DATE:
;       APPLICATION NUMBER:  GB 9518272.1
;       FILING DATE:  08-SEP-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9605550.4
;       FILING DATE:  15-MAR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9607532.0
;       FILING DATE:  11-APR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9609576.5
;       FILING DATE:  08-MAY-1996
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  518 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
US-08-836-620A-18

  Query Match             21.0%;  Score 588.5;  DB 2;  Length 518;
  Best Local Similarity   33.6%;  Pred. No. 3.6e-54;
  Matches  184;  Conservative   68;  Mismatches  212;  Indels   83;  Gaps   19;

Qy          1 MDIKNSPSSLN--------SPSSYNC-----------SQSTLPLE---HGSTYIPSSYVD 48
              |  |:|| ||:          | |  |           |:   | |    |    |    :
Db          5 MSSKDSPGSLDGFAWEDAQKPQSAWCGGRKTEVYATSSRRAPPSEGTRRGGAARPE---E 61

Qy         49 SHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTP---GHLSPLVVHRQL 95
              :  | |     :|  : :|          ||        |  | | |     :|  ||  :
Db         62 AAEEGPPA---APGSLRHS----------GPLGPHACPTAL-PEPQVTSAMSSQVVGTEP 107

Qy         96 SHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCA------- SPVTGPG ----- 141
               :: |||       :  |   | |     |    :  ||              |||
Db        108 LYIKAEPASPDSPKGSSETETEP--PVALAPGPAPTRCLPGHKEEEDGEGAGPGEQGGK 165

Qy        142 ---SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRR 198
                 |       | || | |||||||| ||| ||||||:|||  :| |||:|: | | ||
Db        166 LVLSSLPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGSIEYSCPASNECEITKRRR 225

Qy        199 RSCQACRLRKCYEVGMVKGGSRKERGYRLVRRQRSADEQL-----HCAGKAKRSGGHA 252
               |:|||||  ||  |||:| | | :|  | |   ::|    : |         ||     :||
Db        226 KACQACRFTKCIRVGMLKEGVRLDRVRGGRQKYKRRPEVDPLPFPGPFPAGPLAVAGG   283

Qy        253 PRVRELLLLDALSPEQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLTKLADKELVHMIS 311
              ||     ::|      ||  ||  ||  :  :  |: |       : :|  | |:|:|  ||
Db        284 PRKTAAPVNA----  LVSHLLVVEPEKLYAMPDPAGPDGHLPAVAILCDLFDREIVVTIS 338

Qy        312 WAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVE 371
              ||| |||| |||  || |: :|:| ||||||:|:  ||:    :|  || ||||  :||
Db        339 WAKSIPGFSSISLSDQMSVIQSVWMEVLVIGVAQRSLPLQDELAFAEDLVLDEEGARAA 397

Qy        372 GILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLN 431
              |: |:   ||    | : |:|: :||: :||: | ||    :    :   |  || |
Db        398 GLGELGAALLQLVRRLQALRLEREEYVLLKALALANSDSVHIEDEPRIWSSCEKLLE--  454

Qy        432 AVTDALV WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDL 490
                  :||: :   :::|     :  |   ||: |  :|   :  | : |     :|  :  ||:: |
Db        455 ---EALLEYEAGRAGPGGGAERRRAGRLLLTLPLLRQTAGKVLAHFYGVKLEGKVPMHKL 511

Qy        491 LLEMLNA 497
               |||| |
Db        512 FLEMLEA 518


RESULT 31
US-09-141-000-6
; Sequence 6, Application US/09141000
; Patent No. 6054295
; GENERAL INFORMATION:
;  APPLICANT: Chen, Fang
;  TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR
;  TITLE OF INVENTION: RECEPTOR PROTEINS
;  FILE REFERENCE: 19999Y
;  CURRENT APPLICATION NUMBER: US/09/141,000
;  CURRENT FILING DATE: 1998-08-26
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 418
;   TYPE: PRT
;   ORGANISM: Human
US-09-141-000-6
```

```
  Query Match             20.5%;  Score 576;  DB 3;  Length 418;
  Best Local Similarity   35.6%;  Pred. No. 5.6e-53;
  Matches  152;  Conservative   81;  Mismatches  118;  Indels   76;  Gaps   20;

Qy          17 NCSQSILPLEHGSIYIPSSYVDS-HHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTS 75
               :|| | :  |  |   |:|  || :|  |  :  | |  :||  | :   : |
Db          44 SCS-SFIKTEPSS---PASLTDSVNHHSPGGS--SDASGSYS--STMNGHQNGLDSPPLY 85

Qy          76 PNVLWPTPGHLSPLVVHRQL-----SHLYAFPQKSPWCFARSLFHTLIPVNPEILKRAVSG 130
               |:   |  |   |:   |:|     | :  :||    ||    : :::|
Db          86 PSA--PILGGSGPV---RKLYIDDCSSTIVEDPQTK--CE--YMLNSMP----------- 124

Qy         131 NRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPAINQ 190
                        ||    | || | |||||||| ||| |||||||:|||: :| ||||||:
Db         125 -----------KR---LCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNE 169

Qy         191 CTIDKNRRKSCQACRLRKCYEVGMVKGGSR--RERGGYRLVRPQRSAIPQLHPAPKAPRS 248
               | | | ||||||||||  || :|||:| | |  | | | :  :|:  |:
Db         170 CEITKRRRKSCQACRFMKCLKVGMLKEGVRLDRVRGGRQKYKRRIDAE----------- 217

Qy         249 GGHAPRVP-FLIIDALSP-FQIVETLLEAFPPHV-LISPPSAPFTEASMMMSLTKLADKE 305
                 ::| :  :|:  | |  | ::|  || |||  :  :  |: | ::   : :|  |||:|
Db         218 --NSPYLNPQLVQPAKKPYNKIVSHLLVAEPEKIYAMPDPTVPDSDIKALTTLCDLADRE 275

Qy         306 LVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRD 365
               || :| ||| |||| ||| || :|:| |||:|:--|---||:    :|---| :--| | |
Db         276 LVVIIGWAKHIPGFSTLSLADQMSLLQSAWMEILILGVVYRSLSFEDELVYADDYIMDED 335

Qy         366 EGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNS-SMYPLVTATQDADSSR 424
               : | : |:|:: : :|    ::: :||: :|:: :||: | || ||:      :| :: :
Db         336 QSK-LAGLLDLNNAILQLVKKYKSMKLEKEEFVTLKAIALANSDSMH-----IEDVEAVQ 389

Qy         425 KLAHLLN 431
               ||  :|:
Db         390 KLQDVLH 396


RESULT 32
US-08-764-870-7
; Sequence 7, Application US/08764870
; Patent No. 6236946
;  GENERAL INFORMATION:
;    APPLICANT:  Scanlan, Thomas S
;    APPLICANT:  Baxter, John D
;    APPLICANT:  Fletterick, Robert J
;    APPLICANT:  Wagner, Richard L
;    APPLICANT:  Kushner, Peter J
;    APPLICANT:  Apriletti, James W
;    APPLICANT:  West, Brian
;    TITLE OF INVENTION:  Nuclear Receptor Ligands and Ligand
;    TITLE OF INVENTION:  Binding Domains
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward
;      STREET:  Five Palo Alto Square,  3000 El Camino Real
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,870
;      FILING DATE:  13-DEC-1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/008,540
;      FILING DATE:  13-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/008,543
;      FILING DATE:  13-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/008,606
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakamura, Jackie N
;      REGISTRATION NUMBER:  35,966
;      REFERENCE/DOCKET NUMBER:  UCAL-246/01US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650)843-5000
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  525 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-764-870-7

  Query Match             16.4%;  Score 460.5;  DB 4;  Length 525;
  Best Local Similarity   30.2%;  Pred. No. 2e-40;
  Matches  140;  Conservative   61;  Mismatches  165;  Indels   97;  Gaps   16;

Qy          62 VINLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCFARSLFHTLIPVN- 120
               :::  | ||    :|     ||              :: |  ||   :    :||
Db         128 ISSSMGSPGLPPPAP------PG--------------FSGPVSSPQINSTV---SLPGGG 164

Qy         121 ---RETLKRKVSGNR---CASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                  | :|  | | |   |  |  |||: :    ||:| | :|| ||||:|||||| |||
Db         165 SGPPFDVKPPVIGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCPGCKGFFK 222

Qy         175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVRPQPS 234
               |:|:    | |     ||:|| :|  || || :||   || :   : ||       :| :
Db         223 RTIRKDLTYSCRDNKDCIVDKRQRNRCQYCRYQKCLATGMKREAVQEERQ------QRSKD 276

Qy         235 ADEQLHCAGKAFKRSCGHAPPVPELLLDAISPFQL-VITLLEAFPPHVLI---------- 282
                |   |||  |                    ||:: |  :||||   : :
Db         277 KDGDGECAGGA  ----    -- -  -----PEEMPVDRILEAE---LAVEQKSDQGVEGP 314

Qy         283 --   -SRFSAFTLKVMMMSLTFLADFFLVHMISWAKEIPGFVELSLFDQVRLLESCWMEVL 340
                  | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : :|
Db         315 GGTGGSGSSPNDPVT------NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 368

Qy         341 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQH 399
                |:|:     ||||    :: | | : |:        |:  ||| :|    |: |:::
Db         369 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDK 427

Qy         400 KEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVIDALVWVIAKSGISSQQQSMR 459
                |  |::|:|| |     |         |:     :|     | :    |      ||   |
Db         428 TELGCLRAIILFNPDAKGL--------SNPSEVEVLREKVYASLETYCKQKYPEQQG--R 477

Qy         460 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 502
                | ||: |  :|    | :|||   |     |:   |:||| |
Db         478 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 520


RESULT 33
US-08-980-115-7
; Sequence 7, Application US/08980115
; Patent No. 6266622
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Thomas S.
;  APPLICANT: Baxter, John D.
;  APPLICANT: Fletterick, Robert J.
;  APPLICANT: Wagner, Richard L.
;  APPLICANT: Kushner, Peter J.
;  APPLICANT: Apriletti, James W.
;  APPLICANT: West, Brian L.
;  APPLICANT: Shiau, Andrew K.
;  TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
;  FILE REFERENCE: UCAL-246/02US
;  CURRENT APPLICATION NUMBER: US/08/980,115
```

```
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: 08/764,870
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 60/008,606
;  EARLIER FILING DATE: 1995-12-14
;  EARLIER APPLICATION NUMBER: 60/008,543
;  EARLIER FILING DATE: 1995-12-13
;  EARLIER APPLICATION NUMBER: 60/008,540
;  EARLIER FILING DATE: 1995-12-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 525
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (292)..(523)
;   OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-7

  Query Match             16.4%;  Score 460.5;  DB 4;  Length 525;
  Best Local Similarity   30.2%;  Pred. No. 2e-43;
  Matches  140;  Conservative   61;  Mismatches  165;  Indels   97;  Gaps   16;

Qy          62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN- 120
               :::  | ||    :|      ||              :: |  ||   :     :||
Db         128 ISSSMGSPGLPPPAP-  -  PG-- --------  -FSGPVSSPQINSTV- -SLPGGG 164

Qy         121 -- RETLKPKVSGNP-  -CASPVTGPGSKPDAHPCAVCSDYASGYHYGVWSCEGCKAFFK 174
                 | :|  | | |    |  |  ||| : :    ||:| | :|| ||||:|||||| |||
Db         165 SGPPEDVKPPVLGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFK 222

Qy         175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRMLKHKRQRD 234
               |:|:    | |     ||:|| :|  || || :|    || :   : ||      :| :
Db         223 RTIRKDLTYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEER------QRQKD 276

Qy         235 ADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQL-VLTLLEAEPPHVLI---------- 282
               |    ||| |                ||:: |  :|||    : :
Db         277 KDGDGECAGGA-----------------PEEMPVDRILEAE---LAVEQKSDQGVEGP 314

Qy         283 -------SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
                    | |: | |       :: : |||:  :: |||:|| |  | | ||| || : |
Db         315 GGTGGSGSSPNDPVT -----NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 368

Qy         336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQH 394
                |:|:     ||||    :: |  | : |:       |:  ||| :|    |: |::::
Db         369 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDK 427

Qy         395 KEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR- 454
               |  |::|:|| |     |         |:     :|     | :     | |  ||   |
Db         428 TELGCLRAIILFNPDAKGL- ------SNPSEVEVLREKVYASLETYCKQKYPEQQG  R 477

Qy         455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
               | ||: |  :|    | :|||   |      |:   |:||| |
Db         478 FAKLLLKLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 520


RESULT 34
US-07-952-800-4
; Sequence 4, Application US/07952800
; Patent No. 5403925
;  GENERAL INFORMATION:
;    APPLICANT:  OZATO, KEIKO
;    TITLE OF INVENTION:  A NEW MEMBER OF THE NUCLEAR HORMONE
;    TITLE OF INVENTION:  RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND AND TOWNSEND
;      STREET:  1 MARKET PLAZA, STEUART TOWER, SUITE 2000
;      CITY:  SAN FRANCISCO
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/952,800
;      FILING DATE:  19920928
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Kenneth A
;      REGISTRATION NUMBER:  31,677
;      REFERENCE/DOCKET NUMBER:  15280 21-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-543-9600
;      TELEFAX:  415 543-5043
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  446 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-952-800-4

  Query Match             16.4%;  Score 459;  DB 1;  Length 446;
  Best Local Similarity   29.9%;  Pred. No. 2.2e-40;
  Matches  155;  Conservative   65;  Mismatches  191;  Indels  108;  Gaps   21;

Qy           7 PSSGNSPSSYNCSQSILELE-HGSTYTPSSYVDSHHEYPAMTEYSPAVMNYSIPSNVTNL 65
               | | :| |    || | |    |    ||:         | |    |  :    | :::
Db           8 PRSPDSSSPNPLSQCIRPSSPPGPPLTPSA------PPPPM---PPPPLGSPFPV-ISSS 52

Qy          66 EGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN-   R 121
                | ||    :|      ||              :: |  ||   :     :||
Db          53 MGSPCLPPPAP------PG--------------FSGPVSSPQINSTV---SLPGGGSGPP 89

Qy         122 PTLKPKVSGNP---CASPVTGPGSKPDAHPCAVCSEYASGYHYGVWSCEGCKAFFKRSIQ 178
               | :|  | | |    |  |  ||| :    ||:| | :|| ||||:|||||| ||||:|:
Db          90 EDVKPPVLGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFKRTIR 147

Qy         179 GHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRREPCGYRIVRRQRSADEQ 238
               | |      ||:|| :|  || || :||    || :   : |         ||| |:
Db         148 KDLTYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEF--------RQRGKDKD 199

Qy         239 LHCAGKAKRSGGHAPRVRELLLDALSPEQL-VLTLLEAEPPHVLI         --- -    - 282
               |    :||              :|||: |  :||||    : :
Db         200 ----GDGDGAGG-------------APEEMPVDRILEAE---LAVEQKSDQGVEGPGATG 239

Qy         283 ---SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVL 339
                  | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : |:|
Db         240 GGGSSPNDPVT------NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGWNELL 293

Qy         340 MMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYL 398
               :    ||||     :: |  | : |:       |:  ||| :|    |: |::::    |
Db         294 IASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDKTELG 352

Qy         399 CVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL 458
               |::|:|| |     |         |:     :|     | :     |     ||   | | |
Db         353 CLRAIILFNPDAKGL--------SNPGEVEILREKVYASLETYCKQKYPEQQG  RFAKL 402

Qy         459 LMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
               | :  :|     | :|||   |      |:   |:||| |
Db         403 LKKLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 441


RESULT 35
US-07-952-800-2
; Sequence 2, Application US/07952800
; Patent No. 5403925
;  GENERAL INFORMATION:
;    APPLICANT:  OZATO, KEIKO
;    TITLE OF INVENTION:  A NEW MEMBER OF THE NUCLEAR HORMONE
```

```
;    TITLE OF INVENTION:  RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND AND TOWNSEND
;      STREET:  1 MARKET PLAZA, STEUART TOWER, SUITE 2000
;      CITY:  SAN FRANCISCO
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/952,800
;      FILING DATE:  19920928
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Kenneth A
;      REGISTRATION NUMBER:  31,677
;      REFERENCE/DOCKET NUMBER:  15280-21-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-543-9600
;      TELEFAX:  415-543-5043
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  533 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-952-800-2

  Query Match             16.3%;  Score 458.5;  DB 1;  Length 533;
  Best Local Similarity   30.2%;  Pred. No. 3.4e-40;
  Matches  140;  Conservative   63;  Mismatches  163;  Indels   97;  Gaps   17;

Qy         62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN- 120
              :::  | ||    :|      ||                :: (  ||   :     :||
Db        136 ISSSMGSPGLPPPAP------PG--------------PSGPVSSPQINSTV---SLPGGG 172

Qy        121 ---RETLKPKVSGNR---CASPVTGPGSKPDAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                  | :|  | | |   |  |  |||: :    ||:| | :|| ||||:|||||| |||
Db        173 SGPPEDVKPPVIGVRGLHCPPPPCGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFK 230

Qy        175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRIVPRQRS 234
              |:|:    | |     ||:|| :|  || || :||   || :   : |          |||
Db        231 RTIRKDLTYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEE--------RQRD 282

Qy        235 ADEQLHCAGKAKRSGGHAPPVRELLIDAISPFQL-VETLLFAEPPHVLI---------- 282
               |:     |  : :||             :|||: |  :||||   : :
Db        283 KDKD-   GDGEGAGG-          --- APEEMPVDRILEAE---IAVEQESDQGVEGP 322

Qy        283 -------SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
                    | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : |
Db        323 GGTGGSGSSPNDFVI      NIPQAADKQLELIVEWAKRIPHFSSLPLEDQVILLRAGW 376

Qy        336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQH 394
              |:|:    ||||    :: | | : |:      |:  ||| :|    |: |::::
Db        377 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMPMDK 435

Qy        395 KEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR 454
              |  |:::|:|| |     |        |:     :|     | :    |     |(  |
Db        436 TELGCLRAIILFNPDAKGL--------SNPSEVEVLREKVYASLETYCKQKYPEQQR--R 485

Qy        455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
              | ||: |  :|    | :|||    |      |:   |:|||
Db        486 FAKLLLILPALRSTGLKCLEFLFFFKLIGDTPIDTFLMEMLEA 529


RESULT 36
US-08-336-408B-6
; Sequence 6, Application US/08336408B
```

```
; Patent No. 5723329
;  GENERAL INFORMATION:
;    APPLICANT:  EVANS, RONALD M.
;    APPLICANT:  MANGELSDORF, DAVID J.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 SOUTH FLOWER STREET, SUITE 2000
;      CITY:  LOS ANGELES
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,408B
;      FILING DATE:  08-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/933,453
;      FILING DATE:  21 AUG 1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US91/00399
;      FILING DATE:  22-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REITER, STEPHEN E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9851
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  463 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-336-408B-6

  Query Match             16.1%;  Score 452;  DB 1;  Length 463;
  Best Local Similarity   27.9%;  Pred. No. 1.3e-39;
  Matches  141;  Conservative   72;  Mismatches  215;  Indels   78;  Gaps   15;

Qy          7 PSSLNSPSSYNQSQSILPLEHGSIYIPSSY-VDSHHEYPAMTFYSPAVMNYSIPSNVINL 65
              |:        :  | |: |    |: :|:   :|||  |      :|  :  |       |
Db         12 PTGFGGSPGHTGSTSMSP----SVALPTGKPMDSHPSYTDFPVSAPRIL--SAVGTPLNA 65

Qy         66 EGGFGRQTTSPNVIWPTPCHLS  PLVVHRQLSHLYAEPQKSPWCEAFSLEHTLFVNPET 125
              | | |  ||   : |  | |:  | :      :| | |                    ::
Db         66 LGSPYRVITS--AMGPPSGALAAPPGI------NLVAPPS----------------SQLN 101

Qy        126 LKPKVSGNRCASPV-------------TGPGSKPDAHFCAVCSDYASGYHYGVWSCEGCK 170
              :   ||  :   |:              | |||    | ||:| | :|| ||||:|||||
Db        102 VVNSVSSSSEDIKPLPGLPGTGNMNYPSTSPGS-LVKHICAICGDPSSGKHYGVYSCEGCK 160

Qy        171 AFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRIVP 230
              ||||:|:    | |    | ||| :|  || || :||  :||  :   : ||      |
Db        161 GFFKRTIRKDLTYTCRENKDCSIDKRQKNRCQYCRYQKCLAVGMSREAVQEERQRDREN- 219

Qy        231 PGPSADEQLHCAGKAKRSGGHAPPVRELLIDAISPFQLVETLLFAEPPHVLISPFSAPFT 290
              :  |: :  ||     | :|     | :|:|          |  ||       :   :
Db        220 SRERAESEAECA-----SSSHEDMPVERILEA---------ELAVEPKTESYGDMNVENS 261

Qy        291 EASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDH 350
                  : ::   |||:|  :: |||:|| | :|:| ||| || : | |::|     ||:
```

```
Db        262 TNDPVTNICHAADKQLFTLVEWAKRIPHFSDLTLEDQVTLLRAGWNELLIASFSHRSVSV 321

Qy        351 PGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSS 409
              :: |  | : |        |:  ||| :|    |: :::::   |  |::|::| |
Db        322 QDGILLATGLHVHRSSAHS-RGVGSIFDRVLTELVSKMKDMQMDKSELGCLRAIVLFNPD 380

Qy        410 MYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHAS 469
              |   ::      |:   | | |                  :|  | | ||: |  :|
Db        381 AKGLSNPSEVETLREKVYATLEAYTKQ-- -- ---KYPEQPGRFAKLLLRLPALRSIG 430

Qy        470 NKGMEHLLNMKCKNVVPVYDLLLEML 495
              | :|||   |     |:   |:|||
Db        431 LKCLEHLFFFKLIGDTPIDSFLMEML 456


RESULT 37
PCT US91 00399-6
; Sequence 6, Application PC/TUS9100399
;  GENERAL INFORMATION:
;    APPLICANT:  Mangelsdorf, Dr., David J.
;    APPLICANT:  Evans Dr., Ronald M.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FITCH, EVEN, TABIN & FLANNERY
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/00399
;      FILING DATE:  19910122
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Mr., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  50852
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 552-1411
;      TELEFAX:  (619) 552-0095
;      TELEX:  20 6566 PATLAW CGO
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  463 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT US91 00399 6

  Query Match             16.1%;  Score 452;  DB 5;  Length 463;
  Best Local Similarity   27.9%;  Pred. No. 1.3e-39;
  Matches  141;  Conservative   72;  Mismatches  215;  Indels   78;  Gaps   15;

Qy          7 PSSLNSPSSYNCSQSTLPIFHQSTYIPSSY-VDSHHEYPAMTFYSPAVMNYSIDSNVYNI 65
              |:       :  | |: |    |: :|:   :|||  |       :|  :  |       |
Db         12 PTGFGGSPGHTGSTSMSP---SVALPTGKPMDSHPSYTDTPVSAPRTL--SAVGTPLNA 65

Qy         66 EGGPGRQTTSPNVLWPTPGHLS-PLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRET 124
              | | |  ||   : |  | |: |  :       :| | |                 ::
Db         66 LGSPYRVITS-AMGPPSGALAAPPGI-----NLVAPPS---------- -- --SQLN 101

Qy        124 LKRKVSGNRCASPV-  -  ------ TGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCK 170
                 :  || :    |:            | |||    | ||:| | :|| ||||:||||||

Db        102 VVNSVSSSEDIKPLPGLPGTGNMNYPSPSPGS-LVKHICAICGDRSSGKHYGVYSCEGCK 160

Qy        171 AFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              |||||:|    | |      | ||| :|  || || :||  :|| :   : ||      |
Db        161 GFFKRTIRKDLIYTCRDNKDCLIDKRQRNRCQYCRYQKCLVMGMKREAVQEER    QR 215

Qy        231 EQRSADEQLECAGKAKRSGGHAPEVREILLDALSPEQLVLTLLEAEPPHVLISRPSAPFT 290
              :  |: :  ||     |  |     | :|:|          |  ||        :  :
Db        216 SREKAESEAECA-----SSSHEDMPVERILEA -- --- -ELAVEPKTESYGDMNVENS 261

Qy        291 EASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDH 350
              : ::   |||:|  :: |||:|| | :|:| ||| || : | |:|:     ||:
Db        262 TNDPVTNICHAADKQLFTLVEWAKRIPHFSDLTLEDQVTLLRAGWNELLIASFSHRSVSV 321

Qy        351 PGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSS 409
              :: |  | : |        |:  ||| :|    |: :::::   |  |::|::| |
Db        322 QDGILLATGLHVHRSSAHS-RGVGSIFDRVLTELVSKMKDMQMDKSELGCLRAIVLFNPD 380

Qy        410 MYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHAS 469
              |   ::      |:   | | |                  :|  | | ||: |  :|
Db        381 AKGLSNPSEVETLREKVYATLEAYTKQ-- -------KYPEQPGRFAKLLLRLPALRSIG 430

Qy        470 NKGMEHLLNMKCKNVVPVYDLLLEML 495
              | :|||   |     |:   |:|||
Db        431 LKCLEHLFFFKLIGDTPIDSFLMEML 456


RESULT 38
5223606-6
;Patent No. 5223606
;    APPLICANT:  BLAUDIN DE THE, HUGHES;MARCHIO, AGNES;TIOLLAIS,
;PIERRE;DEJEAN, ANNE
;    TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED
;PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA
;    NUMBER OF SEQUENCES: 11
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/134,130
;      FILING DATE: 17-DEC-1987
;    PRIOR APPLICATION DATA:
;SEQ ID NO:6:
;      LENGTH: 363
5223606-6

  Query Match             16.1%;  Score 451;  DB 6;  Length 363;
  Best Local Similarity   32.4%;  Pred. No. 1.1e-39;
  Matches  105;  Conservative   64;  Mismatches  125;  Indels   30;  Gaps   10;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | :| | ||| |||| :|  || ||||:::| ::|:|   | | :|| |||:| |||||
Db          1 CLICGDEASGCHYGVLTCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRKKNCPACRLRK 60

Qy        209 CYEVGMVKGGSRKERPGGYPLVPRQPSADEQLHCVGKAKPSGCHAPRVRELLLDALSPEQ- 267
              | : |||  | : ::  :  ||  |: |       |  :    |  : :    || |
Db         61 CQQAGMVLGGRKFKK--FNKVRVMRALD------AVALPAPVGTPNESQRI--TFSPSQE 110

Qy        268 ------LVLTLLEAEPPHVLISRPSA-PFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
                   |:  |: :||  :      :  | | :|:: || :| :::|: :: |:| :|||
Db        111 IQLIPPLINLIMSIEPDVIYAGHDNTKPDTSSSLLTSLNQLGERQLLSVVKWSKSLPGFR 170

Qy        321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGKCVEGILEIFD 378
              | : ||: |::  || ::: || |||    |       | ||||||:|:    |      :
Db        171 NLHIDDQITLIQYSWMSLMVFGLGWRSYKHVSGQMLYFAPDLILNEQRMK ESSFYSLCL 229

Qy        379 MLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALV 438
              :      | :|::  :|:|||| ::|||:   ||    :  |  :  : ::    |:
Db        230 TMWQIPQEFVKLQVSQEEFLCMKVLLLINT--IPL  --EGLRSQSQFEEMRSSYIRELI 283

Qy        439 WVIA-  -KSGISSQQQSMRLANLL 459
                 |    |  :|| |:  :|  ||
Db        284 KAIGLRQKGVVSSSQREYQLTKLL 307


RESULT 39
```

```
US-08-764-870-14
; Sequence 14, Application US/08764870
; Patent No. 6236946
;  GENERAL INFORMATION:
;    APPLICANT:  Scanlan, Thomas S
;    APPLICANT:  Baxter, John D
;    APPLICANT:  Fletterick, Robert J
;    APPLICANT:  Wagner, Richard L
;    APPLICANT:  Kushner, Peter J
;    APPLICANT:  Apriletti, James W
;    APPLICANT:  West, Brian
;    TITLE OF INVENTION:  Nuclear Receptor Ligands and Ligand
;    TITLE OF INVENTION:  Binding Domains
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward
;      STREET:  Five Palo Alto Square,  3000 El Camino Real
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,870
;      FILING DATE:  13-DEC-1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/008,540
;      FILING DATE:  13-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/008,543
;      FILING DATE:  13-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/008,606
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakamura, Jackie N
;      REGISTRATION NUMBER:  35,966
;      REFERENCE/DOCKET NUMBER:  UCAL 246/02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650)843-5000
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  933 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-764-870-14

  Query Match             16.1%;  Score 451;  DB 4;  Length 933;
  Best Local Similarity   27.2%;  Pred. No. 5.5e-39;
  Matches  131;  Conservative   81;  Mismatches  163;  Indels  106;  Gaps   18;

Qy          6 SPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHE------YPAMTF-------YSPA 52
              :|    :| :  |   :|| :     :||:   :          |||:         | |
Db        476 APPPCKAPGASGC---LLPRDG----LPSTSASAAAAGAAPALYPAIGLNGLPQLGYQAA 528

Qy         53 VMNYSIPSNVTNLPGGPGRQTTSPNVIWPTPGHISPLVVHPQLSHLYAEPQKSPWCEARS 112
              |:   :                 | |  |   :| |           :|  :||
Db        529 VLKEGL   --------------PQVYPPYLNYLRP----------DSEASQSP------ 555

Qy        113 LEHTLPVNRETIKRKVSGNRCASPVTGPQSKRDAHFCAVCSDYASGYHYGVWSCEGCKAF 172
                  : |:| :|:                     | :| | ||| |||| :|  || |
Db        556 -----QYSFESLPQKI-------------------CLICGDEASGCHYGVLTCGSCKVF 590

Qy        173 FKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREPCGYRLVPRQ 232
              |||:::| :::|:|   | | :|| |||:| |||||||  : |||  | : ::  :  ||
Db        591 FKRAMEGQHNYLCAGRND-CIVDKIRRKNCPACRLRKCQQAGMVLGGRKFKK--FNKVRVV 648

Qy        233 RSADPQI--HCAGKAKPSGGHAPPVPFLLIDALSPPQ-------LVETILEAEPPHVLTS 283
              |: |      |    |   : |          || |       |:  |:  ||  :
Db        649 RALDAVALPQPLGVPNESQALSQRF------TFSPGQDIQLIPPLINLLMSIEPDVIYAG 702

Qy        284 RPSA-PFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMG 342
                :  | | :|:: || :| :::|: :: |:| :|||  | : ||: |::  || ::: |
Db        703 HDNTKPDSSSLLLSLNQLGERQLLSVVKWSKSLPGFRNLHIDDQITLIQYSWMSLMVFG 762

Qy        343 LMWRSIDHPG--KLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCV 400
              | |||  |     | ||||:|:    |       :   :       | :|::  :|:||:
Db        763 LGWRSYKHVSGQMLYFAPDLILNEQRMK-ESSFYSLCLTMWQIPQEFVKLQVSQEEFLCM 821

Qy        401 KAMILLNS-SMYPLVTATQ-DADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL 458
              | ::|||:  :  | : ||  :   |  :  |: |     : :  |  :|| |:  :| |
Db        822 KVLILLNTIPLEGLRSQTQFEEMRSSYIRELIKA-----IGLDQKGVVSSSQRFYQLTKL 876

Qy        459 L 459
              |
Db        877 L 877


RESULT 40
US-08-980-115-14
; Sequence 14, Application US/08980115
; Patent No. 6266622
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Thomas S.
;  APPLICANT: Baxter, John D.
;  APPLICANT: Fletterick, Robert J.
;  APPLICANT: Wagner, Richard L.
;  APPLICANT: Kushner, Peter J.
;  APPLICANT: Apriletti, James W.
;  APPLICANT: West, Brian L.
;  APPLICANT: Shiau, Andrew K.
;  TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
;  FILE REFERENCE: UCAL-246/02US
;  CURRENT APPLICATION NUMBER: US/08/980,115
;  CURRENT FILING DATE: 1997-11-26
;  EARLIER APPLICATION NUMBER: 08/764,870
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 60/008,606
;  EARLIER FILING DATE: 1995-12-14
;  EARLIER APPLICATION NUMBER: 60/008,543
;  EARLIER FILING DATE: 1995-12-13
;  EARLIER APPLICATION NUMBER: 60/008,540
;  EARLIER FILING DATE: 1995-12-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 933
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (659)..(918)
;   OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-14

  Query Match             16.1%;  Score 451;  DB 4;  Length 933;
  Best Local Similarity   27.2%;  Pred. No. 5.5e-39;
  Matches  131;  Conservative   81;  Mismatches  163;  Indels  106;  Gaps   18;

Qy          6 SPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHE------YPAMTF-------YSPA 52
              :|    :| :  |   :|| :     :||:   :          |||:         | |
Db        476 APPPCKAPGASGC---LLPRDG----LPSTSASAAAAGAAPALYPAIGLNGLPQLGYQAA 528

Qy         53 VMNYSIPSNVTNLEGGPGRQTTSPNVIWPTPGHLSPLVVHPQLSHLYAEPQKSPWCEARS 112
              |:   :                 | |  |   :| |           :|  :||
Db        529 VLKEGL-----------------PQVYPPYLNYLRP----------DSEASQSP------ 555

Qy        113 LEHTLPVNRETIKRKVSGNRCASPVTGPQSKRDAHFCAVCSDYASGYHYGVWSCEGCKAF 172
                  : |:| :|:                     | :| | ||| |||| :|  || |
```

```
Db        556 -QYSFESLPQKI- ---------- --------CLICGDEASGCHYGVLTCGSCKVF 590

Qy        174 FKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRPERCGYPLVRRQ 232
              ||||:::| ::|:|   | | :|| |||:| |||||||  : ||| | : ::  :  ||
Db        591 FKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRKCCQAGMVLGGRKFKK--FNKVRVV 648

Qy        233 RSADEQL  HCAGKAKRSGGHAPRVRELLLDALSPEQ------LVLTLLEAEPPHVLIS 283
              |: |       |    |   : |          || |       |:  |:  ||  :
Db        649 RALDAVALPQPLGVPNESQALSQRF------TFSPGQDIQLIPPLINLLMSIEPDVIYAG 702

Qy        284 RPSA-PFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMG 342
                :  | | :|:: || :| :::|: :: |:| :|||  | : ||: |::  || ::: |
Db        703 HDNTKPDTSSSLLTSLNQLGERQLLSVVKWSKSLPGFRNLHIDDQITLIQYSWMSLMVFG 762

Qy        343 LMWRSIDHPG--KLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCV 400
              | ||| |     | |||||:|:    |       :   :       | :|::  :|:||:
Db        763 LGWRSYKHVSGQMLYFAPDLILNEQRMK-ESSFYSLCLTMWQIPQEFVKLQVSQEEFLCM 821

Qy        401 KAMILLNS-SMYPLVTATQ-DADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL 458
              | ::|||:  :  | : || :   |  :  |: |     : :  |  :|| |:  :|  |
Db        822 KVLLLLNTIPLEGLRSQTQFEEMRSSYIRELIKA---  IGLRQKGVVSSSQRFYQLTKL 876

Qy        459 L 459
              |
Db        877 L 877


Search completed: June 16, 2003, 13:39:07
Job time : 28.0717 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 13:15:18 ; Search time 57.7301 Seconds
(without alignments)
981.555 Million cell updates/sec

Title: US-08-826-361A-25
Perfect score: 2805
Sequence: 1 MDIKNSPSSLNSPSSYNQSQ ... PQSPAFDSKSEPQSQNPQSQ 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2805	100.0	530	9	US-10-198-785-2	Sequence 2, Appli
2	1234.5	44.0	595	10	US-09-853-033-2	Sequence 2, Appli
3	1224.5	43.7	595	9	US-10-096-710-1	Sequence 1, Appli
4	1224.5	43.7	595	9	US-10-081-563-2	Sequence 2, Appli
5	1224.5	43.7	595	9	US-10-052-092-9	Sequence 9, Appli
6	1224.5	43.7	595	9	US-10-052-092-13	Sequence 13, Appl
7	1224.5	43.7	595	9	US-10-052-092-14	Sequence 14, Appl
8	1224.5	43.7	595	9	US-09-952-680A-10	Sequence 10, Appl
9	1224.5	43.7	595	10	US-09-933-267A-2	Sequence 2, Appli
10	1219.5	43.5	595	9	US-10-052-092-31	Sequence 31, Appl
11	1219	43.5	595	9	US-10-052-092-30	Sequence 30, Appl
12	1199	42.7	575	10	US-09-893-666A-2	Sequence 2, Appli
13	1190	42.4	239	10	US-09-903-876-2	Sequence 2, Appli
14	1174.5	41.9	701	9	US-10-052-092-12	Sequence 12, Appl
15	1174.5	41.9	701	9	US-10-211-618-1	Sequence 1, Appli
16	960.5	34.2	1099	9	US-10-259-864-4	Sequence 4, Appli
17	922	32.9	1147	9	US-10-259-864-1	Sequence 1, Appli
18	813.5	29.0	910	9	US-09-908-153B-40	Sequence 40, Appl
19	814.5	29.0	922	9	US-09-908-153B-42	Sequence 42, Appl
20	770	27.5	323	10	US-09-747-255-5	Sequence 5, Appli
21	768	27.4	241	10	US-09-905-176-14	Sequence 14, Appl
22	763	27.2	547	9	US-10-052-092-10	Sequence 10, Appl
23	763	27.2	660	10	US-09-853-033-4	Sequence 4, Appli
24	762	27.2	243	10	US-09-903-876-1	Sequence 1, Appli
25	760	27.1	660	10	US-09-853-033-8	Sequence 8, Appli
26	751	26.8	660	10	US-09-853-033-6	Sequence 6, Appli
27	746.5	26.6	264	9	US-10-334-583-1	Sequence 1, Appli
28	745.5	26.6	235	9	US-09-885-827-2	Sequence 2, Appli
29	596.5	21.3	519	9	US-10-221-239-1	Sequence 1, Appli
30	554	19.8	101	10	US-09-864-761-41519	Sequence 41519, A
31	504	18.0	99	10	US-09-864-761-48783	Sequence 48783, A
32	467.5	16.7	689	9	US-10-202-846-2	Sequence 2, Appli
33	458.5	16.3	533	9	US-09-922-226-2	Sequence 2, Appli
34	458.5	16.3	533	9	US-10-098-184-2	Sequence 2, Appli
35	447.5	16.0	601	10	US-09-853-386-92	Sequence 92, Appl
36	447.5	16.0	601	10	US-09-853-386-94	Sequence 94, Appl
37	446	15.9	463	9	US-09-922-226-3	Sequence 3, Appli
38	446	15.9	463	9	US-10-098-184-3	Sequence 3, Appli
39	442.5	15.8	467	10	US-09-965-703-63	Sequence 63, Appl
40	442.5	15.8	598	9	US-10-211-239-2	Sequence 2, Appli
41	442.5	15.8	598	10	US-09-853-386-91	Sequence 91, Appl
42	442.5	15.8	598	10	US-09-853-386-98	Sequence 98, Appl
43	441	15.7	597	10	US-09-853-386-93	Sequence 93, Appl
44	438.5	15.6	777	9	US-10-153-668-226	Sequence 226, App
45	438	15.6	462	9	US-09-922-226-1	Sequence 1, Appli
46	438	15.6	462	9	US-10-098-184-1	Sequence 1, Appli
47	435	15.5	472	10	US-09-965-703-61	Sequence 61, Appl
48	435	15.5	907	9	US-10-008-739A-2	Sequence 2, Appli
49	427.5	15.2	577	10	US-09-853-386-42	Sequence 42, Appl
50	424.5	15.1	1079	9	US-10-021-486B-2	Sequence 2, Appli
51	423	15.1	388	9	US-09-997-267-2	Sequence 2, Appli
52	423.5	15.1	264	9	US-09-933-267-4	Sequence 4, Appli
53	420	15.0	428	10	US-09-965-703-31	Sequence 31, Appl
54	420	15.0	450	9	US-10-245-370-4	Sequence 4, Appli
55	419	14.9	428	10	US-09-965-703-36	Sequence 36, Appl
56	418	14.9	466	9	US-09-909-556-2	Sequence 2, Appli
57	418	14.9	466	10	US-09-909-672-2	Sequence 2, Appli
58	400	14.3	600	10	US-09-853-386-34	Sequence 34, Appl
59	397	14.2	598	10	US-09-853-386-37	Sequence 37, Appl
60	397	14.2	598	10	US-09-853-386-38	Sequence 38, Appl
61	397	14.2	598	10	US-09-853-386-39	Sequence 39, Appl
62	397	14.2	598	10	US-09-853-386-41	Sequence 41, Appl
63	396.5	14.1	598	10	US-09-853-386-40	Sequence 40, Appl
64	394.5	14.1	598	10	US-09-853-386-32	Sequence 32, Appl
65	394.5	14.1	598	10	US-09-853-386-35	Sequence 35, Appl
66	394.5	14.1	598	10	US-09-853-386-36	Sequence 36, Appl
67	394.5	14.1	598	10	US-09-853-386-45	Sequence 45, Appl
68	391	13.9	535	10	US-09-853-386-95	Sequence 95, Appl
69	391	13.9	597	10	US-09-853-386-97	Sequence 97, Appl
70	385	13.7	643	10	US-09-853-386-70	Sequence 70, Appl
71	378.5	13.5	627	10	US-09-853-386-67	Sequence 67, Appl
72	378.5	13.5	627	12	US-10-005-169-2	Sequence 2, Appli
73	378	13.5	73	9	US-09-879-746-3	Sequence 3, Appli
74	377.5	13.5	628	10	US-09-853-386-66	Sequence 66, Appl
75	377.5	13.5	628	10	US-09-853-386-68	Sequence 68, Appl
76	377.5	13.5	628	10	US-09-853-386-73	Sequence 73, Appl
77	377.5	13.5	628	12	US-10-005-169-4	Sequence 4, Appli
78	375	13.4	626	10	US-09-853-386-65	Sequence 65, Appl
79	375	13.4	626	10	US-09-853-386-96	Sequence 96, Appl
80	374.5	13.4	461	10	US-09-731-557A-2	Sequence 2, Appli
81	374	13.3	626	10	US-09-853-386-64	Sequence 64, Appl
82	370.5	13.2	454	9	US-09-814-604-3	Sequence 3, Appli
83	370.5	13.2	454	10	US-09-797-727-2	Sequence 2, Appli
84	367.5	13.1	625	10	US-09-853-386-63	Sequence 63, Appl
85	367	13.1	448	10	US-09-797-727-4	Sequence 4, Appli
86	367	13.1	462	9	US-09-814-604-1	Sequence 1, Appli
87	367	13.1	462	10	US-09-797-727-3	Sequence 3, Appli
88	366.5	13.1	1089	9	US-10-259-864-2	Sequence 2, Appli
89	366.5	13.1	1090	9	US-10-259-864-6	Sequence 6, Appli
90	364	13.0	448	9	US-09-814-604-2	Sequence 2, Appli
91	360	12.8	423	10	US-09-919-497-86	Sequence 86, Appl
92	359.5	12.7	461	12	US-10-013-823-3	Sequence 3, Appli

93	352.5	12.6	513	10	US-09-965-703-59	Sequence 59, Appl
94	352.5	12.6	746	9	US-09-042-488B-7	Sequence 7, Appli
95	351.5	12.5	625	10	US-09-965-703-16	Sequence 16, Appl
96	351.5	12.5	764	9	US-10-300-757-20	Sequence 20, Appl
97	349.5	12.5	395	9	US-10-151-542A-21	Sequence 21, Appl
98	346.5	12.4	446	12	US-10-013-823-2	Sequence 2, Appli
99	344.5	12.3	446	10	US-09-909-446-2	Sequence 2, Appli
100	344.5	12.3	446	10	US-09-909-325-2	Sequence 2, Appli
101	344.5	12.3	446	10	US-09-909-326-2	Sequence 2, Appli
102	343.5	12.2	746	9	US-10-300-757-18	Sequence 18, Appl
103	342	12.2	1041	9	US-09-042-488B-9	Sequence 9, Appli
104	340.5	12.1	546	9	US-10-295-870-2	Sequence 2, Appli
105	339.5	12.1	746	9	US-09-042-488B-5	Sequence 5, Appli
106	333	11.9	61	9	US-09-814-604-21	Sequence 21, Appl
107	333	11.9	61	9	US-09-922-226-7	Sequence 7, Appli
108	333	11.9	61	9	US-09-922-226-160	Sequence 160, App
109	330	11.8	367	9	US-10-151-542A-2	Sequence 2, Appli
110	324.5	11.5	440	10	US-09-965-703-71	Sequence 71, Appl
111	317.5	11.3	412	10	US-09-965-703-11	Sequence 11, Appl
112	317.5	11.3	412	10	US-09-965-703-12	Sequence 12, Appl
113	315	11.2	60	10	US-09-864-761-40315	Sequence 40315, A
114	312.5	11.1	606	10	US-09-952-559-3	Sequence 3, Appli
115	307.5	11.0	469	9	US-10-155-479-2	Sequence 2, Appli
116	300	10.7	476	9	US-10-188-721-1	Sequence 1, Appli
117	298	10.6	348	10	US-09-760-364-1	Sequence 1, Appli
118	297	10.6	359	10	US-09-760-364-9	Sequence 9, Appli
119	292.5	10.4	477	9	US-10-182-232-3	Sequence 3, Appli
120	290	10.4	264	9	US-10-043-487-254	Sequence 254, App

ALIGNMENTS

```
RESULT 1
US-10-198-785-2
; Sequence 2, Application US/10198785
; Publication No. US20030022224A1
; GENERAL INFORMATION:
;  APPLICANT: OLYMPUS OPTICAL CO., LTD.
;  TITLE OF INVENTION: Method of detecting binding reaction between protein and test sub
;  FILE REFERENCE: ZEI 02S0538
;  CURRENT APPLICATION NUMBER: US/10/198,785
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: JP/2001 220444
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: JP/2001-221963
;  PRIOR FILING DATE: 2001-07-23
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 530
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-198-785-2

  Query Match             100.0%;  Score 2805;  DB 9;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-237;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHPQLSHLYAEPQKSPWCEARSLEHLLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHPQLSHLYAEPQKSPWCEARSLEHLLPVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCEAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVRRQRSADEQLH 240

Qy        241 QAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 PSSPRLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PSSPRLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530


RESULT 2
US-09-853-033-2
; Sequence 2, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-853-033-2

  Query Match             44.0%;  Score 1233.5;  DB 10;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 2.2e-99;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      | :|||   | :|: ||   | :: | | | | | :  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ---- ------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :              : || :|   |  ||| | | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||       |    :  |: |    :      |     | ::  :||||:|
Db        134 NEPSGYIVREAGPPAFYRP---NSDNRRQGGREERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRREERGGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR------ - - - -EL 258
              :| | |::| | |::: :|  |:      |: :  | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||:||::
Db        306 LALSLTADQMVSALLDAEPP ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364
```

```
Qy          317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
                ||||:|:| ||| ||| |:|:||:||:|||::|||||:|||:|:|||::|||||||:||
Db          365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLILDRNQGKCVEGMVEI 424

Qy          377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
                |||||||:|||| : || :|:|:|:|:|||| :|  :::| :  :    :  :|: :||
Db          425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy          436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                 |: ::||:|:: ||| ||| ||::|||:|| |||||||| :|||||||:|||||||
Db          485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy          496 NAHVL 500
                :|| |
Db          545 DAHRL 549


RESULT 3
US-10-096-710-1
; Sequence 1, Application US/10096710
; Patent No. US20020164581A1
; GENERAL INFORMATION:
;  APPLICANT: Chang, Chawnshang
;  TITLE OF INVENTION: METHODS FOR IDENTIFYING COMPOUNDS THAT
;  TITLE OF INVENTION:  DO NOT ATTENUATE THE PROTECTIVE EFFECTS OF ESTROGEN
;  FILE REFERENCE: 21108.001002
;  CURRENT APPLICATION NUMBER: US/10/096,710
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/275,229
;  PRIOR FILING DATE: 2001-03-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:/note =
;   OTHER INFORMATION: synthetic construct
US-10-096-710-1

  Query Match             43.7%;  Score 1224.5;  DB 9;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 1.4e-98;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy            5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
                |    || |      | :|||   | :|: ||   | :: |  |  | : |  |  :
Db           21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy           61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                  |  ||| :                 : || :|   |  ||| |  | |    | |
Db           75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy          102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
                : | :   ||        |    .  |.  |    .      |     | ::  :|||:|
Db          134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy          155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
                ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db          191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy          215 VKCGSRRFRCGYRLVPPQRSADEQLHCAGKAKRSGGHAPRVR----------------EL 258
                :| | |:|| | |::: :|  |:      |: :   | |  :|
Db          251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy          259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
                |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|::
Db          306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy          317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
                ||||:|:| ||| |||  |:|:||:||:|||:.|| ||.|||.|.|||..||||||..||
Db          365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy          377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
                |||||||:|||| : || :|:|:|:|:|||| :|  :::| :  :    :  :|: :||
Db          425 FDMILATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy          436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                 |: ::||:|:: ||| ||| ||::|||:|| |||||||| :|||||||:|||||||
Db          485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy          496 NAHVL 500
                :|| |
Db          545 DAHRL 549


RESULT 4
US-10-081-563-2
; Sequence 2, Application US/10081563
; Publication No. US20020187495A1
; GENERAL INFORMATION:
;  APPLICANT: Wake Forest University
;  APPLICANT:  Herrington, David M.
;  APPLICANT:  Howard, Timothy D.
;  APPLICANT:  Hawkins, Gregory A.
;  APPLICANT:  Meyers, Deborah A.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS OF ESTROGEN RECEPTOR ALPHA ASSOCIA
;  TITLE OF INVENTION:  FAVORABLE HDL CHOLESTEROL RESPONSE TO HORMONE REPLACEMENT THE
;  FILE REFERENCE: 9151.15
;  CURRENT APPLICATION NUMBER: US/10/081,563
;  CURRENT FILING DATE:  2002-02-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-081-563-2

  Query Match             43.7%;  Score 1224.5;  DB 9;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 1.4e-98;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy            5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
                |    || |      | :|||   | :|: ||   | :: |  |  | : |  |  :
Db           21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy           61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                  |  ||| :                 : || :|   |  ||| |  | |    | |
Db           75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy          102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
                : | :   ||        |    .  |.  |    .      |     | ::  :|||:|
Db          134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy          155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
                ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db          191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy          215 VKCGSRRFRCGYRLVPPQRSADEQLHCAGKAKRSGGHAPRVR----------------EL 258
                :| | |:|| | |::: :|  |:      |: :   | |  :|
Db          251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy          259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
                |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|::
Db          306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy          317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
                ||||:|:| ||| |||  |:|:||:||:|||:.|| ||.|||.|.|||..||||||..||
Db          365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy          377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
                |||||||:|||| : || :|:|:|:|:|||| :|  :::| :  :    :  :|: :||
Db          425 FDMILATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy          436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
```

```
                 |: ::||:|:: ||| ||| ||::|||:|| |||||||| :|||||||:|||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 5
US-10-052-092-9
; Sequence 9, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Human
US-10-052-092-9

  Query Match             43.7%;  Score 1224.5;  DB 9;  Length 595;
  Best Local Similarity   48.8%;  Pred. No. 1.4e-98;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      | :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP-----QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSFAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||       |    :  |:  |    :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||:|||| ||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR------------ --EL 258
              :| | |:|| | |::: :|  |:      |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:|:|||::|| ||:||||:|||::|||||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||||:|  :::| :  :     :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 6
US-10-052-092-13
; Sequence 13, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: human
US-10-052-092-13

  Query Match             43.7%;  Score 1224.5;  DB 9;  Length 595;
  Best Local Similarity   48.8%;  Pred. No. 1.4e-98;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      | :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSFAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||       |    :  |:  |    :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||:|||| ||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR------------ -- EL 258
              :| | |:|| | |::: :|  |:      |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:|:|||::|| ||:||||:|||::|||||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||||:|  :::| :  :     :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549
```

```
RESULT 7
US-10-052-092-14
; Sequence 14, Application US/10052092
; Publication No. US20020177778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Human
US-10-052-092-14

  Query Match             43.7%;  Score 1224.5;  DB 9;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 1.4e-98;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMIFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: | |  | | : | |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ----------------TTSPNVLWPTPGHISP-LVVHRQLSHIYAF 101
                | |  ||| :               : || :|   | ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--FAPSIEHTIPVNRFTLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||      |    :  |:  |    :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRPQGGRFRLASTNDKGSMAMESAKFTRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSRRFRCGYRLVRRQRSADPQLHCAGKAKRSGGHAPRVR----------------FL 258
              :| | |::| | |::: :|  |:     |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRPDD-----GEGPGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPFQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:||| :| |   |: ||:|||||  || |||:|||||:|:::
Db        306 LALSLIADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              |||||:|:| ||| |||  |:|:||:|||:|||::|| ||:|||:|:|||::||||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSKKLAHLLNAVTD 435
              |||||||:|||| : || :|:|:|:|:|||| :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::|||:|:: |||  ||| ||::|||:|| |||||||:|||||||:||||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 8
US-09-952-680A-10
; Sequence 10, Application US/09952680A
; Publication No. US20030087349A1
; GENERAL INFORMATION:
;  APPLICANT: [illegible]son, Matt
;  APPLICANT:  Epstein, David
;  APPLICANT:  Hamaguchi, Ne. US20030087349Aluko
;  TITLE OF INVENTION: Target Activated Biosensor and Methods of Using Same
;  FILE REFERENCE: 23239-501
;  CURRENT APPLICATION NUMBER: US/09/952,680A
;  CURRENT FILING DATE:  2001-09-13
;  PRIOR APPLICATION NUMBER: 60/232,454
;  PRIOR FILING DATE: 2000-09-13
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-952-680A-10

  Query Match             43.7%;  Score 1224.5;  DB 9;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 1.4e-98;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMIFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: | |  | | : | |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ----------------TTSPNVIWPTPGHISP-LVVHRQLSHIYAF 101
                | |  ||| :               : || :|   | ||| |  | |    | |
Db         75 Q-PGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--FAPSIEHTIPVNRFTLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||      |    :  |:  |    :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRPQGGRFRLASTNDKGSMAMESAKFTRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKCGSRRFRCGYRLVRRQRSADPQLHCAGKAKRSGGHAPRVR----------------FL 258
              :| | |::| | |::: :|  |:     |: :   | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRPDD-----GEGPGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPFQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:||| :| |   |: ||:|||||  || |||:|||||:|:::
Db        306 LALSLIADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              |||||:|:| ||| |||  |:|:||:|||:|||::|| ||:|||:|:|||::||||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSKKLAHLLNAVTD 435
              |||||||:|||| : || :|:|:|:|:|||| :|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::|||:|:: |||  ||| ||::|||:|| |||||||:|||||||:||||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 9
US-09-933-267A-2
; Sequence 2, Application US/09933267A
; Patent No. US20020123095A1
; GENERAL INFORMATION:
;  APPLICANT: Kalush, Francis et al.
;  TITLE OF INVENTION: Estrogen receptor alpha variants and
```

```
;  TITLE OF INVENTION:  methods of detection thereof
;  FILE REFERENCE: CL000258C14
;  CURRENT APPLICATION NUMBER: US/09/933,267A
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 60/160626
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: 60/184756
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: 09/692414
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 09/768184
;  PRIOR FILING DATE: 2001-01-24
;  PRIOR APPLICATION NUMBER: 09/804076
;  PRIOR FILING DATE: 2001-03-13
;  PRIOR APPLICATION NUMBER: 09/826414
;  PRIOR FILING DATE: 2001-04-05
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: HUMAN
US-09-933-267A-2

  Query Match             43.7%;  Score 1224.5;  DB 10;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1.4e-98;
  Matches  263;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |       |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 QITGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 134

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
              : | :   ||       |    :  |:  |    :      |     | ::  :||||:|
Db        135 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:|||:||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRERGGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR----------    --EL 258
              :| | |::| | |::: :|  |:     |: :  | |  :|
Db        251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:||:|||::|| ||:|||:|:|||::||||||::||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||||  :|  :::| :  :     :  :|: :||
Db        425 FDMLLATSSKFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
               |: ::||:|:: ||  ||| ||::|||:|| |||||||| :|||||||:||||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :|| |
Db        545 DAHRL 549


RESULT 10
US-10-052-092-31
; Sequence 31, Application US/10052092
; Publication No. US20030027778A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: pig
US-10-052-092-31

  Query Match             43.5%;  Score 1219.5;  DB 9;  Length 595;
  Best Local Similarity   46.0%;  Pred. No. 4.8e-98;
  Matches  268;  Conservative   89;  Mismatches  126;  Indels   99;  Gaps   19;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSSYVDSHHEYPAMTFYSPAVMNY------ 56
              |    || |       |  :|||   | :|: ||                 ||| ||
Db         21 NELEPLNRP------QLKIPLERPLGEVYVDSS---- --  ---  KPAVYNYPEGAAY 60

Qy         57 --------SIP-SNVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP- 88
                      | |    : |  ||| :                 : || ||   |  |||
Db         61 DFNAAAAASAPVYGQQSGLAYGPGSEAAAFGANGLGGFQPLNSVSPSPLVLLHPPPQLSPF 120

Qy         89 LVVHRQLSHLYAEPQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG---- 141
              |  | |    | | : | :   ||       |    :  |:  |    :  :  |
Db        121 LHPHGQQVPYYLENEPSGYAVPEAGPPAFYRP---NSDNRRQGGRERLASISDKGSMAME 177

Qy        142 SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSC 201
              | ::  :||||:|||||||||||||||||||||||||||||:|||||||||||||||||
Db        178 SAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSC 237

Qy        202 QACRLRKCYEVGMVKGGSRRERGGYRLVRRQRSADEQLHCAGKAK---------RSGG-- 250
              |||||||||||||:| | |::| | |::: :|  |:     |: :          ||
Db        238 QACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDD-  ---GEGRNEAVPPGDMRSANLW 292

Qy        251 -------HAPRVRELLLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKL 301
                    |  :   :|  :|: :|::  |||||||  :: |   |: } :|||||  || |
Db        293 PSPLLIKHTKKNSPVL--SLTADQMISALLEAEPP-IIYSEYDPTRPLSEASMMGLLTNL 349

Qy        302 ADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLV 361
              ||:|||||:|||::|||::||| ||| |||  |:|:||:||:|||::|||||::||||:|
Db        350 ADRELVHMINWAKRVPGFLDLSLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLL 409

Qy        362 LDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT QDA 420
              |||::||||||::|||||||||:|||| : || :|:::|:|::|||||  :|  :::| :
Db        410 LDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSL 469

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :    :  :|: :|| |: ::||:|:: |||  ||| ||::|||:|| |||||||| |||
Db        470 EEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHRRLAQLLLILSHFRHMSNKGMEHLYNMK 529

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKE 522
              ||||||:|||||||:|| |    ::: |    | ||    :
Db        530 CKNVVPLYDLLLEMLDAHRLH-APTNLGG---PPPEDMSQSQ 567


RESULT 11
US-10-052-092-30
; Sequence 30, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
```

```
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: hamster
US-10-052,092-30

  Query Match             43.5%;  Score 1219;  DB 9;  Length 595;
  Best Local Similarity   46.7%;  Pred. No. 4.2e-98;
  Matches  262;  Conservative   93;  Mismatches  134;  Indels   72;  Gaps   15.

Qy         23 LPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNY----------SIPSNV---TNLEGG 68
              :|||     :   ||||              ||: ||          : |: |   | :  |
Db         33 MPLEPA---LSFVYVDSS---------KPAMFNYPEGAAYFFNAATAPAPVYGQTGIAYG 80

Qy         69 PGRQTT----------------SPN---VIWPTPGHISP IVVHPQISHIYAFPQFSAWT 108
                | : |                ||:   :| | |  ||| |  | |    | | : | :
Db         81 SGSEATAFGSNSLGIFPQLNSVSPSPLMLIPPPPPQLSPFLHPHGQQVPYYLFNFPSAYA 140

Qy        109 EARSLEHTLFVNRETLKFKVSGNFCASPVTGPGSF-----RDAHFCAVGFPYASGYHYGV 163
                |          :  |: ||    |  :  ||       ::  :||||:||||||||
Db        141 -VRDSGPPAFYPSNSDNPPQSGPPRLSSSSFKGSMAMESVKFTPYGAVGNDYASGYHFGV 199

Qy        164 WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGVRKDR 223
              ||||||||||||||||||||||:|||||||||||||||||||||||||||||||:| | |::|
Db        200 WSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDR 259

Qy        224 CGYRLVPRQRSADE-----QLHCACKARSG--------CHAPPVPEELLLDALSPDQLVL 270
               | |::: :|  |:      :  :|  : :         |  :    |  :|: :|:|
Db        260 RGGRMLKHKRQRDDDEGRNDMGPSGDMRAANLWPSPLVIKHTKKNSPAL  SLTADQMVS 317

Qy        271 TLLEAEPPHVLIS-RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVR 329
              ||:|||| :     || ||:|||||  || |||:|||||||:|||:|||  :|:| |||
Db        318 ALLDAEPPLIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRVPGFGDLNLHDQVH 377

Qy        330 LLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRE 389
              ||| |:|:||:||:|||::|||||:|||:|:|||::|||||||::|||||||||:|||
Db        378 LLECAWLEILMIGLIWRSMEHPGKLIFAPNLLLDRNQGKCVEGMVEIFDMLLATSARFRM 437

Qy        390 LKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALVWVIAKSGISS 448
              : || :|:|:|::|||||:|  :||  : :  :  :|: :|| |: ::||:|::
Db        438 MDLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTL 497

Qy        449 QQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSII 508
              |||  ||| ||::|||:|| |||||||| ||||||||:|| |     : ::
Db        498 QQQHRRLAQLLILLSHIRHMSNKGMEHLYNMKCKNVVPFYDLLLEMLDAHRLH---TPVS 554

Qy        509 GSPGSPAPDSKSKPGSQNPQS 529
                  || | |:|:  : |  |
Db        555 RMGVSPPEPSQSQLITTNSTS 575


RESULT 12
US-09-893-666A-2
; Sequence 2, Application US/09893666A
; Patent No. US20020129386A1
; GENERAL INFORMATION:
;  APPLICANT: YAMASHITA, ICHIRO
;  TITLE OF INVENTION: High estrogen-sensitive medaka fish
;  FILE REFERENCE: 210217US-620-7249-0
;  CURRENT APPLICATION NUMBER: US/09/893,666A
;  CURRENT FILING DATE:  2002-02-12
;  PRIOR APPLICATION NUMBER: JP 2000-247729
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 575
;   TYPE: PRT
;   ORGANISM: Oryzias latipes
US-09-893-666A-2

  Query Match             42.7%;  Score 1199;  DB 10;  Length 575;
  Best Local Similarity   48.1%;  Pred. No. 2.2e-96;
  Matches  265;  Conservative   81;  Mismatches  129;  Indels   76;  Gaps   16;

Qy         36 YVDSHHEY----PAMT---------FYSPAVMNYSIPS--NVTNLEGGPGRQTTSPNVLW 80
              :::  ::|    || |         :|| :     ||  :: :|  ||    ||| |
Db         17 FLEGTYDYAAPNPAIIPLYSQSSIGYYSAPLEINGPPSEGSLQSLGSGP---ISPLVFV 72

Qy         81 PTPGHLSPLVVHPQLSHLYAFPQKSPWCEAPSLEHTLPVNPETLKPKVSGNR---CASPV 137
              |:   ||| :     || | |  :|  :          ::   :    |:|   |:
Db         73 PSSPRLSFFM-  HPPSHHYLETTSTFVYRSS         HQGASREDQCGSREDTCSIGE 122

Qy        138 TGPGS-------KRDAHFCAVGSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
               | |:          :|  |||||||||||||||||||||||||||||||||||||||||||||
Db        123 SGACAGAGSTFMAKDTPFCAVGSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQ 182

Qy        191 CTIDKNRRKSCQACRLRKCYEVGMVKGGVRKDR--------------------AD 236
              ||||:|||||||||||||||||||:| | |::|    |:::|| :
Db        183 CTIDKNRRKSCQACRLRKCYEVGMMKGGVRKDR--IPILPPDKPPRTGVGDGDKVVKGQEH 240

Qy        237 EQLECACKAKKE--GGHAPEVPEELLLDALSPEQLVLELLEAEPPHVLISRP--SAPFTEA 292
              : :|  |: : |  ||       | : :: |||::| |  |||| :| ||  | |:||
Db        241 KIVRYLGKRKSSLDGGGGGGGGRLSVISIFPEQVLLLAQGALEPLHLCSMQPELSPYTFV 299

Qy        293 SMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG 352
              :||  || :||||||||:||||:|||::||| ||| |||| (:||||:||:|||  ||
Db        300 IMMILLISMADKELVHMIAWAKKLPGFLQLSLHDQVLLLESSWLEVLMIGLIWRSIEHPG 359

Qy        353 KLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYP 412
              ||||| |||||||:|| |||: |||||||||| |||| ||| :|:|:|||||||  :
Db        360 KLIFAQDLILDRNEGKCVEGMIEIFDMLLAISARFRVLKLKHEEFVCLKAIILLNSGAFS 419

Qy        413 LVIATQD-ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNK 471
              | | :   :|  :  :|: :|||:  |:|||  :|:| | | ||:|||||:|| |||
Db        420 FCTGTMEPLHNSAAVQSMLDTITDALIHYISQSGYLAQEQARRQAQLLILLSHIRHMSNK 479

Qy        472 GMEHLLNMKCKNVVPVYDLLLEMLNAHVL---RGCKS-----------SIIGSPGSPAE 516
              ||||| :||||| ||:|||||||:|| |    |  :|           | |   :|
Db        480 GMEHLYSMKCKNKVPLYDLLLEMLDAHRLHHPVPAPQSLSQVDRQPPSISSGGGGIAPGS 539

Qy        517 DSKSKPGSQNP 527
              | |:   ::|
Db        540 ISASRGRIESP 550


RESULT 13
US-09-903-876-2
; Sequence 2, Application US/09903876
; Patent No. US20020072587A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Zhang-Bao
;  APPLICANT:  Somers, William
;  APPLICANT:  Akopian, Tatos
;  APPLICANT:  Hsiao, Chu-Lai
;  APPLICANT:  Unwalla, Rayomond
;  TITLE OF INVENTION: CRYSTAL STRUCTURE OF ESTROGEN RECEPTOR B COMPLEX
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 37174/9
;  CURRENT APPLICATION NUMBER: US/09/903,876
;  CURRENT FILING DATE:  2001-07-11
;  PRIOR APPLICATION NUMBER: US 60/217,834
;  PRIOR FILING DATE: 2000-07-12
;  NUMBER OF SEQ ID NOS: 2
```

```
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 239
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-903-876-2

  Query Match             42.4%;  Score 1190;  DB 10;  Length 239;
  Best Local Similarity   99.2%;  Pred. No. 3.9e-96;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy        260 LDALSPEQLVLTLLEAEPPHVLISRPSAPETEASMMMSLTKLADKELVHMISWAKKIPGF 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LDALSPEQLVLTLLEAEPPHVLISRPSAPETEASMMMSLTKLADKELVHMISWAKKIPGF 60

Qy        320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 120

Qy        380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVW 439
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        121 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTA-QDADSSRKLAHLLNAVTDALVW 179

Qy        440 VIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHV 499
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        180 VIAKSGISSQQQSMRLANLLM-LSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHV 238

Qy        500 L 500
              |
Db        239 L 239


RESULT 14
US-10-052-092-12
; Sequence 12, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Bopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 701
;   TYPE: PRT
;   ORGANISM: human
US-10-052-092-12

  Query Match             41.9%;  Score 1174.5;  DB 9;  Length 701;
  Best Local Similarity   41.2%;  Pred. No. 4.1e-94;
  Matches  268;  Conservative   87;  Mismatches  125;  Indels  171;  Gaps   16;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTEYSPAVMNYSIPS 60
              |   || |       | :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------ QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ- - - - - - - - - - -TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                  : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHILPVNRELLKRKVSGNRCASPVTGPG- - - SKRDAHFCAVCSD 154
              : | :  ||       |   :  |: |   :       |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRRERGGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR- - - - - - - - - EL 258
              :| | |:|| | |::: :|  |:      |: :  | | :|
Db        251 MKGGIRKDRRGGRMLKHKRQMDD-----GEGMGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPETEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              |||||:|| ||| |||  |:|:||:||:|||:|:|| ||:|||:|:|||:|||||:||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT------QD---- ------ 419
              |||||||:|||| : || :|:::|:|::|||| :|  :::|       :|
Db        425 FDMLLATSSRFRMMNLQGFEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        420 --------- ------ -  - -------- --      -- -- --- ---ADSSR --- 424
                                                                | |||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRL 655

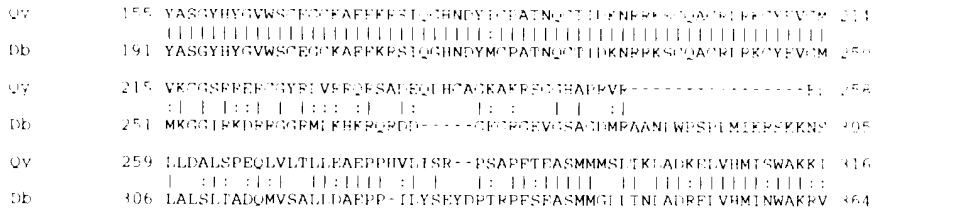

RESULT 15
US-10-211-613-1
; Sequence 1, Application US/10211613
; Publication No. US20030099641A1
; GENERAL INFORMATION:
;  APPLICANT: CELL SIGNALING TECHNOLOGY, INC.
;  APPLICANT:  SMITH, Bradley
;  APPLICANT:  CROSBY, Katherine
;  APPLICANT:  WU, Jiong
;  TITLE OF INVENTION: Monoclonal Antibodies Specific for Phosphorylated Estrogen Rec
;  TITLE OF INVENTION:  (Ser 118) and Uses Thereof
;  FILE REFERENCE: CST 202
;  CURRENT APPLICATION NUMBER: US/10/211,613
;  CURRENT FILING DATE:  2002-08-02
;  PRIOR APPLICATION NUMBER: US 60/310,066
;  PRIOR FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 701
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-211-613-1

  Query Match             41.9%;  Score 1174.5;  DB 9;  Length 701;
  Best Local Similarity   41.2%;  Pred. No. 4.1e-94;
  Matches  268;  Conservative   87;  Mismatches  125;  Indels  171;  Gaps   16;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTEYSPAVMNYSIPS 60
              |   || |       | :|||   | :|: ||   | :: |  |  | : |  |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

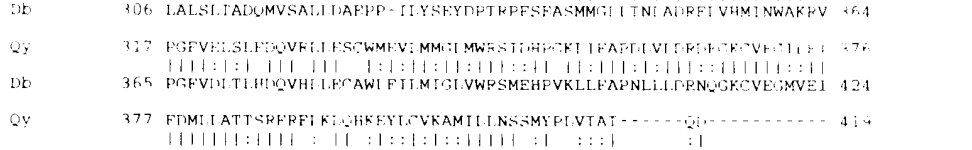
Qy         61 NVTNLEGGPGRQ- - - - - - - - - - -TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                  : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

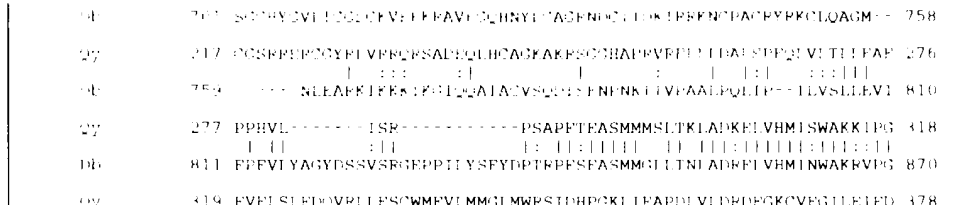
Qy        102 PQKSPWC--EARSLEHILPVNRELLKRKVSGNRCASPVTGPG- - - SKRDAHFCAVCSD 154
              : | :  ||       |   :  |: |   :       |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190
```



RESULT 16
US-10-259-864-4
; Sequence 4, Application US/10259864
; Publication No. US20030077645A1
; GENERAL INFORMATION:
; APPLICANT: Hager, Gordon L
; TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfa
; TITLE OF INVENTION: Receptor Ligands, and Methods and Kits For Detecting and Charact
; TITLE OF INVENTION: Ligands
; FILE REFERENCE: 24799
; CURRENT APPLICATION NUMBER: US/10/259,864
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/325,178
; PRIOR FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1099
; TYPE: PRT
; ORGANISM: Rat/human Chimera
US-10-259-864-4

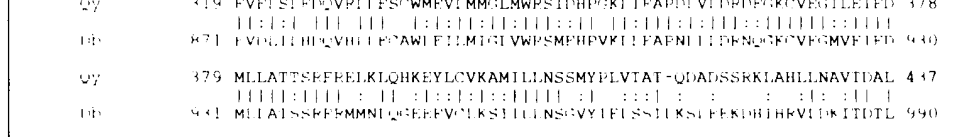
Query Match 34.2%; Score 960.5; DB 9; Length 1099;
Best Local Similarity 40.1%; Pred. No. 4.2e-75;
Matches 218; Conservative 95; Mismatches 159; Indels 71; Gaps 15;





RESULT 17
US-10-259-864-1
; Sequence 1, Application US/10259864
; Publication No. US20030077645A1
; GENERAL INFORMATION:
; APPLICANT: Hager, Gordon L
; TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfa
; TITLE OF INVENTION: Receptor Ligands, and Methods and Kits For Detecting and Char
; TITLE OF INVENTION: Ligands
; FILE REFERENCE: 24799
; CURRENT APPLICATION NUMBER: US/10/259,864
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/325,178
; PRIOR FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Rat/human chimera
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1)..()
; OTHER INFORMATION: Chimeric Protein
US-10-259-864-1

Query Match 32.9%; Score 922; DB 9; Length 1147;
Best Local Similarity 36.5%; Pred. No. 1.1e-71;
Matches 216; Conservative 95; Mismatches 161; Indels 120; Gaps 15;



```
Qy        217 CGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE 276
                  |  :::   :|        |     :   |  |:|   :::|||
Db        758   - --NLEARKTKKKTKGTQQATAGVSQDTSENPNKTIVPAALPQLTP--TLVSLLEVT 809

Qy        277 PPHVLIS                  RPSAPETEASMMMSLTK---------- -          300
              | || :                   |: ||:|||||  ||
Db        810 EPEVLYAGYDSSSRGFPQTLYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVH 879

Qy        301                              - ---LADKELVHMISWAKKIPGFVELSLFDQVR 329
                                           |||:|||||||:|||::|||||:|:| |||
Db        870 KAHQETFPALCQLGKYTTNNSSEQRVSLDTDLADRELVHMINWAKRVPGFVDLTLHDQVH 929

Qy        330 LLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRE 389
              ||| |:|:|||:|:|||::|| ||:|||:|:|||::|||||:|||||||:|||||||:|||| 
Db        930 LLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRM 989

Qy        390 LKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALVWVIAKSGISS 448
              : (( :|::|:|::|||||  :|  :::| :  :    :  :|: :|| |: ::||:|::
Db        990 MNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTL 1049

Qy        449 QQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
              |||  ||| ||::|||:|| |||||||| :|||||||||:|||||||:|| |
Db       1050 QQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRL 1101


RESULT 18
US-09-908-153B-40
; Sequence 40, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beerli, Roger
;  APPLICANT:  Schopter, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40
;  LENGTH: 910
;  TYPE: PRT
;  ORGANISM: Unknown
;  FEATURE:
;  OTHER INFORMATION: Synthesized
US-09-908-153B-40

  Query Match             29.0%;  Score 813.5;  DB 9;  Length 910;
  Best Local Similarity   39.8%;  Pred. No. 2.4e-62;
  Matches  196;  Conservative   83;  Mismatches  129;  Indels   85;  Gaps   17;

Qy         90 VVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFC 149
              :|  | :|   :| | | |  :|   :  :|  :| :  |   :|  :|  |  |  :
Db         32 LVRHQRTHTGEKPYKCPEC-GKSFSQSGDLRRH--QRTHTGEK--PYKCPECGKS---F 82

Qy        150 AVCSDYA--SGYHYG--VWSCEGC-KAF-----  -  FKRSTQGHNDYICPAINQCTIDKN 196
              : | | |     | |   ::|  | ||:|          :|  |   | ||   :|
Db         83 SDCRDLARHQRTHTGEKPYACPECGKSFSQSSHLVRHQRTHTGEKPYKCP--EC ---  144

Qy        197 RRKSCQACR- --- -------LRKCYEVGMVEGCSRRERCGYRLVRRQRSADEQLHCAG 243
              ||   ||                 || | |  |  || :    :||| ||:
Db        145 -GKSFSQSGDLRRHQRTHTGEKPYKCPECG--KSFSRSD---KLVRHQRT        H 179

Qy        244 KAKRSGGHAPRVRELLLD-- -- ---  -------- -------ALSPEQLVLTLLEAEPP 278
              |:: | | |  |:                                :|: :|:|  ||:|||
Db        180 TGKKTSQQACRRNEMGASGDMRAANLWPSPLVIKHTKKNSPALSLTADQMVSALLDAEPP 239

Qy        279 HVLIS-RPSAPETEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWME 337
              :     || ||:|||||  || |||:|||||:|||::||| :|:| ||| ||| |:|
Db        240 MIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRVPGFGDLNLHDQVHLLECAWLE 299

Qy        338 VLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEY 397
              :|:||:|||::|||||:|||:|:|||::|||||:|||||||:|||||||:|||| : || :|:
Db        300 LIMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEF 359

Qy        398 LCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLA 456
              :|:|::|||| :|  :::| :  :    :  :|: :|| |: ::||:|::  ||| |||
Db        360 VCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHRRLA 419

Qy        457 NLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAE 516
              ||::|||:|| |||||||| ||||||||:|||||||:||  |    | :      |::
Db        420 QLLLILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEMLDAHRLHAPASRMGVPPEEPSQ 479

Qy        517 DSKSKEGSQNPQS 529
              :  |: |
Db        480 TQLATTSSTSAHS 492


RESULT 19
US-09-908-153B-42
; Sequence 42, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beerli, Roger
;  APPLICANT:  Schopter, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42
;  LENGTH: 922
;  TYPE: PRT
;  ORGANISM: Unknown
;  FEATURE:
;  OTHER INFORMATION: Synthesized
US-09-908-153B-42

  Query Match             29.0%;  Score 813.5;  DB 9;  Length 922;
  Best Local Similarity   39.8%;  Pred. No. 2.4e-62;
  Matches  196;  Conservative   83;  Mismatches  129;  Indels   85;  Gaps   17;

Qy         90 VVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFC 149
              :|  | :|   :| | | |  :|   :  :|  :| :  |   :|  :|  |  |  :
Db         32 LVRHQRTHTGEKPYKCPEC-GKSFSQSGDLRRH--QRTHTGEK--PYKCPECGKS---F 82

Qy        150 AVCSDYA--SGYHYG--VWSCEGC-KAF-----  -  FKRSTQGHNDYICPAINQCTIDKN 196
              : | | |     | |   ::|  | ||:|          :|  |   | ||   :|
Db         83 SDCRDLARHQRTHTGEKPYACPECGKSFSQSSHLVRHQRTHTGEKPYKCP--EC ---  144

Qy        197 RRKSCQACR- --- -------LRKCYEVGMVEGCSRRERCGYRLVRRQRSADEQLHCAG 243
              ||   ||                 || | |  |  || :    :||| ||:
Db        145 -GKSFSQSGDLRRHQRTHTGEKPYKCPECG--KSFSRSD---KLVRHQRT        H 179

Qy        244 KAKRSGGHAPRVRELLLD-- -- ---  -------- -------ALSPEQLVLTLLEAEPP 278
              |:: | | |  |:                                :|: :|:|  ||:|||
Db        180 TGKKTSQQACRRNEMGASGDMRAANLWPSPLVIKHTKKNSPALSLTADQMVSALLDAEPP 239

Qy        279 HVLIS-RPSAPETEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWME 337
              :     || ||:|||||  || |||:|||||:|||::||| :|:| ||| ||| |:|
Db        240 MIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRVPGFGDLNLHDQVHLLECAWLE 299

Qy        338 VLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEY 397
              :|:||:|||::|||||:|||:|:|||::|||||:|||||||:|||||||:|||| : || :|:
Db        300 LIMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEF 359
```

```
Qy          498 LCVKAMILLNSSMYPLVTAL-QDADSSPKLAHLLNAVTDALVWVIAKSGISSQQQSMRLA 556
                ::|:|::|||||:|  :::| :  :    :  :|: :|| |: ::|||:|:: ||| |||
Db          360 VCLKSILLLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHRRLA 419

Qy          557 NLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAE 616
                ||::|||:|| |||||||| ||||||||:|||||||:|| |   | :      ::
Db          420 QLLLILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEMLDAHRLHAPASRMGVPPEEPSQ 479

Qy          617 DSKSKEGSQNPQS 629
                :  | :  |
Db          480 TQLATTSSTSAHS 492


RESULT 20
US-09-737-255-5
; Sequence 5, Application US/09737255
; Patent No. US20020073441A1
; GENERAL INFORMATION:
;  APPLICANT: Rehemtulla, Alnawaz
;  APPLICANT:  Ross, Brian D.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
;  TITLE OF INVENTION:  PROTEOLYTIC ACTIVITY
;  FILE REFERENCE: 11203-003001
;  CURRENT APPLICATION NUMBER: US/09/737,255
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-737-255-5

  Query Match             27.5%;  Score 770;  DB 10;  Length 323;
  Best Local Similarity   54.4%;  Pred. No. 3.4e-59;
  Matches  147;  Conservative   60;  Mismatches   61;  Indels    2;  Gaps    2;

Qy          262 ALSPEQLVLTLLEAEPPHVLISR-PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
                :|: :|:|  |||:|||| :|  :     || |||:|||||  || |||:|||||:|:|::|||
Db           45 SLTADQMVSALLDAEPPMILYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRVPGFG 104

Qy          321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
                :|:| ||| ||| |:|.||.|:|:|..||||:|||.|.|||..||||:|||:|:|||||
Db          105 DLNIHDQVHLLECAWLEILMIGLVWRSMEHPGKLIFAPNLLLDRNQGKCVEGMVEIFDML 164

Qy          381 LATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAL-QDADSSPKLAHLLNAVTDALVW 439
                |||:|||| : || :|:|:|:|:|||| :|  :::| :  :    :  :|: :|| |:
Db          165 LATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIH 224

Qy          440 VIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHV 499
                ::|:|:: |||  ||| ||:::|:|| |||||||| ||||||||:|:|||||:||
Db          225 IMAKAGLTLQQQHRRLAQLLILLSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEMLDAHR 284

Qy          500 LRGCKSSITGSECSPAEDSKSKEGSQNPQS 529
                |    | :     |::  :   | : |
Db          285 LHAPASRMGVPPEEPSQTQLATTSSTSAHS 314


RESULT 21
US-09-905-176-14
; Sequence 14, Application US/09905176
; Patent No. US20020150906A1
; GENERAL INFORMATION:
;  APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
;  APPLICANT:  Debe, Derek A.
;  TITLE OF INVENTION: METHOD FOR DETERMINING THREE DIMENSIONAL PROTEIN STRUCTURE FROM P
;  TITLE OF INVENTION:  PROTEIN SEQUENCE
;  FILE REFERENCE: 265/297
;  CURRENT APPLICATION NUMBER: US/09/905,176
;  CURRENT FILING DATE:  2002-04-05
;  PRIOR APPLICATION NUMBER: US 60/218,016
;  PRIOR FILING DATE: 2000-07-12
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
;   LENGTH: 241
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-905-176-14

  Query Match             27.4%;  Score 768;  DB 10;  Length 241;
  Best Local Similarity   59.8%;  Pred. No. 3.8e-59;
  Matches  144;  Conservative   56;  Mismatches   47;  Indels    4;  Gaps    4;

Qy          263 LSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
                |: :|:|  |||:|||| :| |   |: |||:|||||  || |||:|||||:|:|::|||
Db            1 LTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFV 59

Qy          321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
                :|:| ||| |||  |:|.|||:|||:|||:|||||:|||:|:|||||:|||||||||||
Db           60 DLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDML 119

Qy          381 LATISRFRELKLQHKEYLCVKAMILLNSSMYPLVTAL-QDADSSKKLAHLLNAVTDALVW 439
                |||:|||| : || :|:|:|:|:|||| :|  :::| :  :    :  :|: :|| |:
Db          120 LATSSPFPMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIH 179

Qy          440 VIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHV 499
                ::|:|:: |||  ||| ||:::|:|| |||||||| ||||||||:|:|||||:||
Db          180 IMAKAGLTLQQQHRRLAQLLILLSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHR 239

Qy          500 L 500
                !
Db          240 L 240


RESULT 22
US-10-052-092-10
; Sequence 10, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapy
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 547
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Estrogen Receptor Ligand Binding Domain Fusion
US-10-052-092-10

  Query Match             27.2%;  Score 761;  DB 9;  Length 547;
  Best Local Similarity   45.4%;  Pred. No. 3e-58;
  Matches  164;  Conservative   66;  Mismatches   83;  Indels   48;  Gaps    7;

Qy          190 HNDYICPATNQCTIDKNRRKSCQAGFRRLGIYEVGMVKCGSPPPP----------GGY 239
                || ||       | || :       ::|   | | |:     ::           |
Db          149 HNVYI-------TADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY 201

Qy          240 RLVRRQRSADEQ-------LHCAGKAKRSGGHAPEVR          ---  ELLLCA 282
                :  | |           |     |:  || |  :|                |  :
Db          202 LSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK 241

Qy          283 LSELEVLVALGLVYRSFLSSTLEVTKSLALKGGALQYVHMKGAEHVVGRD 332
                |: :|:|  |||:|||| :| |   |: |||:|||||  || |||:|||||:|:|::|||
Db          242 LSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK 281
```

```
Db        262 LTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFV 320

Qy        421 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 480
              :|:| ||| |||  |:|:||:||:|||::|| ||:|||:|:|||::|||||||::|||||
Db        321 DLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDML 380

Qy        481 LATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALVW 539
              |||:|||| : || :|::|:|::|||| :|  :::| :  :     :  :|: :|| |:
Db        381 LATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIH 440

Qy        540 VIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHV 599
              ::|:|:: |||  ||| |||::|||:|| |||||||| :|||||||||:||||||||:||
Db        441 LMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHR 500

Qy        600 L 600
              |
Db        501 L 501


RESULT 23
US-09-853-033-4
; Sequence 4, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;  LENGTH: 660
;  TYPE: PRT
;  ORGANISM: Artificial sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;  OTHER INFORMATION: Homosapiens-Bacteriophage P1
US-09-853-033-4

  Query Match             27.2%;  Score 763;  DB 10;  Length 660;
  Best Local Similarity   59.1%;  Pred. No. 3.9e-58;
  Matches  143;  Conservative   57;  Mismatches   38;  Indels    4;  Gaps    3;

Qy        262 ALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGF 319
              :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||::|||
Db        374 SLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGF 432

Qy        320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
              |:|:| ||| |||  |:|:||:||:|||::||| :|||:|:|||::|||||||::|||||
Db        433 VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDM 492

Qy        380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVIAT-QDADSSRKLAHLLNAVTDALV 438
              |||:|||| : || :|::|:|::|||| :|  :::| :  :     :  :|: :|| |:
Db        493 LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLI 552

Qy        439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
              ::|:|:: |||  ||| |||::|||:|| ||||||| :|||||||||:||||||||:||
Db        553 HLMAKAGLTLQQQHQRLAQLLILLSHIRHMSNKRMEHLYSMKCKNVVPLYDLLLEMLDAH 612

Qy        499 VL 500
              |
Db        613 RL 614


RESULT 24
US-09-903-876-1
; Sequence 1, Application US/09903876
; Patent No. US20020072587A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Zhanq-Bao
;  APPLICANT:  Somers, William
;  APPLICANT:  Akopian, Tatos
;  APPLICANT:  Hsiao, Chu-Lai
;  APPLICANT:  Unwalla, Rayomond
;  TITLE OF INVENTION: CRYSTAL STRUCTURE OF ESTROGEN RECEPTOR-B COMPLEX
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 37174/9
;  CURRENT APPLICATION NUMBER: US/09/903,876
;  CURRENT FILING DATE:  2001-07-11
;  PRIOR APPLICATION NUMBER: US 60/217,844
;  PRIOR FILING DATE: 2000-07-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;  LENGTH: 243
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-903-876-1

  Query Match             27.2%;  Score 762;  DB 10;  Length 243;
  Best Local Similarity   59.1%;  Pred. No. 1.1e-58;
  Matches  143;  Conservative   57;  Mismatches   38;  Indels    4;  Gaps    3;

Qy        262 ALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGF 319
              :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||::|||
Db          3 SLTADQMVSALLDAEPP ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGF 61

Qy        320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
              |:|:| ||| |||  |:|:||:||:|||::||| :|||:|:|||::|||||||::|||||
Db         62 VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDM 121

Qy        380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVIAT-QDADSSRKLAHLLNAVTDALV 438
              |||:|||| : || :|::|:|::|||| :|  :::| :  :     :  :|: :|| |:
Db        122 LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLI 181

Qy        439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
              ::|:|:: |||  ||| |||::|||:|| ||||||| :|||||||||:||||||||:||
Db        182 HLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAH 241

Qy        499 VL 500
              |
Db        242 RL 243


RESULT 25
US-09-853-033-8
; Sequence 8, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;  LENGTH: 660
;  TYPE: PRT
;  ORGANISM: Artificial sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;  OTHER INFORMATION: Homosapiens-Bacteriophage P1
US-09-853-033-8

  Query Match             27.1%;  Score 760;  DB 10;  Length 660;
  Best Local Similarity   58.7%;  Pred. No. 7.2e-58;
```

```
  Matches  142;  Conservative   57;  Mismatches   40;  Indels    4;  Gaps    4;

Qy         262 ALSPEQLVLILLEAEPPHVLISR--PSAPFTEASMMMSLIKLADKELVHMISWAKKIPGF 319
               :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||::|||
Db         374 SLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGF 432

Qy         320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
               |:|:| ||| |||  |:|:||:|||:|||:|||:|:|||::|||||::|||||
Db         433 VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDM 492

Qy         380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALV 438
               ||||:|||| : || :|:|:|:|::||||| :|  :::| :  :    :  :|: :|| |:
Db         493 LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLI 552

Qy         439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
               ::|:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||  :||
Db         553 HLMAKAGLTLQQQHQRLAQLLILLSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEAADAH 612

Qy         499 VL 500
               |
Db         613 RL 614


RESULT 26
US-09-853-033-6
; Sequence 6, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 06691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 660
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;   OTHER INFORMATION: Homosapiens Bacteriophage P1
US-09-853-033-6

  Query Match             26.8%;  Score 751;  DB 10;  Length 660;
  Best Local Similarity   58.4%;  Pred. No. 4.4e-57;
  Matches  141;  Conservative   57;  Mismatches   40;  Indels    4;  Gaps    3;

Qy         262 ALSPEQLVLILLEAEPPHVLISR--PSAPFTEASMMMSLIKLADKELVHMISWAKKIPGF 319
               :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|||::|||
Db         374 SLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGF 432

Qy         320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
               |:|:| ||| |||  |:|:||:|||:|||:|||:|:|||::|||||::|||||
Db         433 VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEIFDM 492

Qy         380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALV 438
               ||||:|||| : || :|:|:|:|::||||| :|  :::| :  :    :  :|: :|| |:
Db         493 LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLI 552

Qy         439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
               ::|:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||  :||
Db         553 HLMAKAGLTLQQQHQRLAQLLILLSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEAADAH 612

Qy         499 VL 500
               |
Db         613 RL 614


RESULT 27
US-10-334-583-1
; Sequence 1, Application US/10334583
; Publication No. US20030104484A1
; GENERAL INFORMATION:
;  APPLICANT: Cleaver, Brian
;  APPLICANT:  Green, Mike L.
;  TITLE OF INVENTION: Materials and Methods for Detection and Quantitation of an Ana
;  FILE REFERENCE: ELI-101XC1
;  CURRENT APPLICATION NUMBER: US/10/334,583
;  CURRENT FILING DATE:  2002-12-31
;  PRIOR APPLICATION NUMBER: US/09/660,979
;  PRIOR FILING DATE: 2000-09-13
;  PRIOR APPLICATION NUMBER: 60/153,627
;  PRIOR FILING DATE: 1999-09-13
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 264
;   TYPE: PRT
;   ORGANISM: Equus
US-10-334-583-1

  Query Match             26.6%;  Score 746.5;  DB 9;  Length 264;
  Best Local Similarity   59.5%;  Pred. No. 2.9e-57;
  Matches  144;  Conservative   53;  Mismatches   40;  Indels    5;  Gaps    4;

Qy         262 ALSPEQLVLILLEAEPPHVLISRPSA--PFTEASMMMSLIKLADKELVHMISWAKKIPGF 319
               :|: ||::  ||:|||| || |   |  || ||||  || |||:|||||:|||::|||
Db           9 SLTAEQMISALLDAEPP-VLYSEYDATRPFNEASMMGLLTNLADRELVHMINWAKRVPGF  67

Qy         320 VELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
               |:||| ||| |||  |:|:||:|||:|||:|||:|:|||::|||||::|||||
Db          68 VDLSLHDQVHLLECAWLEILMICLVWPSMEHPGKLLFAPNLILDRNQGKCVEGMVEIFDM 127

Qy         380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALV 438
               ||||:|| | : || :|:|:|:|::||||| :|  :::| :  :    :  :|: :|| |:
Db         128 LLATSSRLRMMNLQGEEPVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKMTDTLI 187

Qy         439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
                ::|:|:: ||   ||| ||::|||:|| |||||||| :|||||||:|||||  :||
Db         188 HLMAKAGLTLQQH-RRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAH 246

Qy         499 VL 500
               |
Db         247 RL 248


RESULT 28
US-09-885-827-2
; Sequence 2, Application US/09885827
; Patent No. US20020173445A1
; GENERAL INFORMATION:
;  APPLICANT: Salvati, Mark
;  APPLICANT:  Attar, Ricardo M
;  APPLICANT:  Gottardis, Marco M
;  APPLICANT:  Krystek Jr., Stanley R
;  APPLICANT:  Sack, John S
;  TITLE OF INVENTION: SELECTIVE ANDROGEN RECEPTOR MODULATORS AND METHODS FOR
;  TITLE OF INVENTION:  THEIR IDENTIFICATION, DESIGN AND USE
;  FILE REFERENCE: LD0250(NP)
;  CURRENT APPLICATION NUMBER: US/09/885,827
;  CURRENT FILING DATE:  2001-06-20
;  PRIOR APPLICATION NUMBER: 60/284,438
;  PRIOR FILING DATE: 2001-04-18
;  PRIOR APPLICATION NUMBER: 60/214,392
;  PRIOR FILING DATE: 2000-06-28
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 235
;   TYPE: PRT
;   ORGANISM: Homo sapiens
```

```
US 09 885 827-2

  Query Match              26.6%;  Score 745.5;  DB 9;  Length 235;
  Best Local Similarity    59.1%;  Pred. No. 4e-57;
  Matches  140;  Conservative   53;  Mismatches   37;  Indels    7;  Gaps    3;

Qy        262 ALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVE 321
              :|: :|:|  ||:|||| :| |     |:|||||  || |||:||||||:|||::||||:
Db          5 SLTADQMVSALLDAEPP ILYSE---FSEASMMGLLTNLADRELVHMINWAKRVPGFVD 59

Qy        322 LSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLL 381
              |:| ||| ||| |:|:||:||:|||::|||||:|||:|:|||::|||||||::|||||||
Db         60 LTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL 119

Qy        382 ATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVI 441
              ||:|||| : || ::::|:|::|||||| :|      :  :     :|: :|| |: ::
Db        120 ATSSKFRMMNLQGEEFVCLKSIILLNSGVYTF- TLKSLEEKDHIHRVLDKITDTLIHLM 177

Qy        442 AKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
              ||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||:||
Db        178 AKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAH 234


RESULT 29
US 10 211 239 1
; Sequence 1, Application US/10211239
; Publication No. US20030103965A1
; GENERAL INFORMATION:
;  APPLICANT: Jung, Birgit
;  APPLICANT:  Kraut, No. US20030103965Albert
;  APPLICANT:  Mueller, Stefan
;  TITLE OF INVENTION: Method for Identifying Substances Which Positively
;  TITLE OF INVENTION:  Influence Inflammatory Conditions
;  FILE REFERENCE: 0652.234000
;  CURRENT APPLICATION NUMBER: US/10/211,239
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/415,775
;  PRIOR FILING DATE: 2001-08-30
;  PRIOR APPLICATION NUMBER: EP 01119003.0
;  PRIOR FILING DATE: 2001-08-06
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;  LENGTH: 519
;  TYPE: PRT
;  ORGANISM: Homo Sapiens
US 10 211 239 1

  Query Match              21.4%;  Score 596.5;  DB 9;  Length 519;
  Best Local Similarity    33.0%;  Pred. No. 1e-43;
  Matches  182;  Conservative   70;  Mismatches  206;  Indels   93;  Gaps   19;

Qy          1 MDIKNSPSSLN         SPSSYNC------ --  -SQSILPLE---HGSIYIPSSYVD 48
              |  |:|| ||:          | | |            |:   | |    |    |    :
Db          5 MSSKDSPQSLDGRAWFDAQKPQSAWCGPKTPVYATSSPPAPPSPGTPRGSAAEEF  -P 61

Qy         49 SHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTP---GHLSPLVVHRQL 105
              :  | |      :|  : :|           ||      | | | |      :|  ||  :
Db         62 AAEEGPPA   APGSLRHS           GFLGPHACPTAL-PEPQVTSAMSSQVVGIEP 107

Qy         96 SHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCA---   ---SPVTGPG------ 141
              :: |||      :  |   | |    |    :  ||              |||
Db        108 LYIKAEPASPDSPKGSSETETEP--PVALAPGPAPTRCLPCHKEEEDGEGAGPGEQGGGK 165

Qy        142 -  SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKESIQGHNDYICPATNQCTIDKNRR 198
                 |      | || | ||||||||| ||| |||||||:||   :| |||:|:| | | ||
Db        166 LVLSSLPKRLCLVCGDVASGYHYGVASCEACKAFFKRTIQGSIEYSCPASNECEITKRRR 225

Qy        199 KSCQACRLRKCYEVGMVKCGSR--RERCGYRLVRKQRSADE ---------QLHCAGKAK 246
              |:|||||  ||  |||:| | |  | | | :  :|:   |            |  ||  :
Db        226 KACQACRFTKCLRVGMLKEGVRLDRVRGGRQKYKRRPEVDPLPFPGPFPAGPLAVAGGPR 285

Qy        247 RSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKEL 306
              ::       |  ||:  : |]:|      | |       |: |      : :|  | |:|:
Db        286 KTAPVNALVSHLLV -VEPEKLY--  --AMPD--    -PAGPDGHLPAVATLCDLEDREI 342

Qy        307 VHMISWAKKIPGFVELSLEDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDE 366
              |  |||| ||||  ||| ||: :|:| |||||:|:  ||:    :| || ||||| :|
Db        343 VVTISWAKSIPGFSSLSLSDQMSVLQSVWMEVLVLGVAQRSLPLQDELAFAEDLVLD EE 401

Qy        367 GKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKL 426
              |    |: |:   ||    | : |:|: :||: :||: | ||    :    :|||::  :|
Db        402 GARAAGLGELGAALLQLVRRLQALRLEREEYVLLKALALANSDSVHI-  --EDAEAVEQL 457

Qy        427 AHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVP 486
                |:      | :   ::|     :  |   ||: |  :|  : | : |   :|  ||
Db        448 REALHEA--LLEYEAGRAGPGGGAERRAGRLLLTLPLLRQTAGKVLAHFYGVKLEGKVP 505

Qy        487 VYDLLLEMLNA 497
              :: | |||| |
Db        506 MHKLFLEMLEA 516


RESULT 30
US-09-864-761-41519
; Sequence 41519, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41519
```

```
;   LENGTH: 101
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL   4.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   6
;   OTHER INFORMATION: EST_HUMAN HIT: BF510841.1, EVALUE 5.00e-12
;   OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 8.00e-58
US-09-864-761-41519

  Query Match             19.8%;  Score 554;  DB 10;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-41;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          20 QSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVL 79
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 QSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVL 60

Qy          80 WPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
               |||||||||||||||||||||||||||||||||||||||||
Db          61 WPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 101


RESULT 31
US-09-864-761-48783
; Sequence 48783, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,866
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48783
;   LENGTH: 99
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL   0.46
;   OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 5.00e-52
;   OTHER INFORMATION: EST_HUMAN HIT: AI073549.1, EVALUE 2.00e-11
US-09-864-761-48783

  Query Match             18.0%;  Score 504;  DB 10;  Length 99;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-36;
  Matches   99;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         219 SRREPCGYRLVPPQPSADEQLHCAGKAKRSGGHAPPVRELLLDALSPEQLVLTLLEAEPP 278
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 SRREPCGYRLVPPQPSADEQLHCAGKAKRSGGHAPPVRELLLDALSPEQLVLTLLEAEPP 60

Qy         279 HVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIP 317
               |||||||||||||||||||||||||||||||||||||||
Db          61 HVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIP 99


RESULT 42
US-10-202-846-2
; Sequence 2, Application US/10202846
; Publication No. US20030083487A1
; GENERAL INFORMATION:
;  APPLICANT: NAGAHAMA, Yoshitaka
;  APPLICANT:  IKEUCHI, Toshitaka
;  APPLICANT:  KOBAYASHI, Toru
;  APPLICANT:  TODO, Takashi
;  TITLE OF INVENTION: PROGESTOGEN RECEPTOR POLYPEPTIDES, TRANSGENIC CELLS IN WHICH G
;  TITLE OF INVENTION:  ENCODING SAID POLYPEPTIDES ARE INTRODUCED AND METHOD FOR DETE
;  TITLE OF INVENTION:  FOR PROGESTOGENS BY USING SAID TRANSGENIC CELLS
;  FILE REFERENCE: 113343
;  CURRENT APPLICATION NUMBER: US/10/202,846
;  CURRENT FILING DATE:  2002-07-26
;  PRIOR APPLICATION NUMBER: JP 2001/235,725
;  PRIOR FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
;   LENGTH: 689
;   TYPE: PRT
;   ORGANISM: Anguilla japonica
US-10-202-846-2

  Query Match             16.7%;  Score 467.5;  DB 9;  Length 689;
  Best Local Similarity   27.3%;  Pred. No. 3e-32;
  Matches  148;  Conservative   83;  Mismatches  208;  Indels  103;  Gaps   20;

Qy           4 KNSPSSLNSPSSYNC------SQSILPLEHGSIYI-PSSYVD ------------SHHE 42
               | | | ::| |  ||        |   :|  | |  ||  :                :||
Db         145 KASLSGVSSSSVGNCKFIKDEQDSSSSMEPQSPYFHPSGNIITSNSSYGTCEEDSAIHHP 204

Qy          43 YPAMTFYSPAVMNYSIPSNVINLEGGPGRQTTSPNVLWPTP ---------GHLSPL 89
                 | |:       ||           : |     :| |              | ||
Db         205 PHMFTDYNRTTALPLIP-----------EITEDQPSFPYIPVGEVVANSCLTGYQQRSPQ 252

Qy          90 VVHR------QLSHLYAEPQKSPWCEARSLEHTLPVNRFLTKRFVSGNRCA--------S 135
                 |       :||   : |:   ||:: |          |    | |          |
Db         253 NSLRFKSELCKLSLPTSSPFSQSWCQSTGLSEDQHFETGYLPPCEIRNICEIHNSLKSHS 312
```

```
Qy          146 PVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDK 205
                |  |::    || :| | ||| |||| :|  || |:|||:::|| :|:|    | | :||
Db          313 VYMGMLSQK---FCLICGDEASGCHYGVLTCGSCKVFYKRAVEGHQNYLCAGRNDCTVDK 369

Qy          206 NRRKSCQACRLRKCYEVGMVKCGSRRERCG---------YRLVRRQRSADEQLHCACKAK 256
                 |||:| |||||||||: ||   | : ::                ||    |  :    |:|
Db          370 IRRKNCPACRLRKCYQAGMTLGGRKMKKLSALKVIGLLTQSLAVKSPLGASYE----GQAL 425

Qy          257 KSGGHAFKVRELLLLDALSPEQLVLTLLEAEPPHVLISRPSA- PFTEASMMMSLTKLADK 315
                :    | ||||     :|:  :|::||   |  : |    |  | |    :: ||  | ::
Db          426 ATLPSMPMVREL- -QFTPQ -ILSILENIEPETVYSGYDATQPETPHLLLNSLNGLCER 480

Qy          316 ELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK--LIFAPDLVL 373
                :|: :: |:| :|||  | : || |:  :|  ||  :::  | |||  :  :  | ||||||:|
Db          481 QLLWIVRWSKSLPGFRSLHINDQMTLIQYSWMSLMVFSLGWRSFQNVTREELYFAPDLIL 540

Qy          374 DRDEGKCVEGILEIFDMLLA--TTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAFQD 431
                  :: :       | |: :|       |  |::  :|:||:| ::|||:   ||
Db          541 GEEKMR---NSPISDLCMAMQTIPQAFDNLQVTKEEFLCMKVLLLLNTI--VPLEGLRSQ 594

Qy          432 ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNM 491
                |         :  :| |:   : : |: :  | |  :|  |:  :    |     :|:
Db          595 AQFDEMRHGYIRELIKAI  QLTERGVMASSQ -RFYHLTKLMDAMHELVRK-- --VNL 645

Qy          492 KC 493
                |
Db          646 YC 647


RESULT 33
US-09-922-226-2
; Sequence 2, Application US/09922226
; Publication No. US20030077664A1
; GENERAL INFORMATION:
;  APPLICANT: Zhao, Yi
;  APPLICANT:  Thacher, Scott M.
;  APPLICANT:  Xiao, Jia Bao
;  APPLICANT:  Kusari, Jyotirmoy
;  APPLICANT:  Chandraratna, Roshantha A.
;  TITLE OF INVENTION: Methods of Screening For Compounds That
;  TITLE OF INVENTION:  Modulate Hormone Receptor Activity
;  FILE REFERENCE: P AR 4681
;  CURRENT APPLICATION NUMBER: US/09/922,226
;  CURRENT FILING DATE:  2002-01-09
;  PRIOR APPLICATION NUMBER: US 60/284,797
;  PRIOR FILING DATE: 2001-04-18
;  NUMBER OF SEQ ID NOS: 191
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 533
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-09-922-226-2

  Query Match             16.3%;  Score 458.5;  DB 9;  Length 533;
  Best Local Similarity   40.2%;  Pred. No. 1.3e-41;
  Matches  140;  Conservative   63;  Mismatches  163;  Indels   97;  Gaps   17;

Qy           62 VTNLEGGPGRQTTSPNVILWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN- 120
                :::  | ||    :|     ||                  :: |  ||   :    :||
Db          146 ISSSMGSPGLPPPAP------PG-------------FSGPVSSPQINSTV---SLPGGG 172

Qy          121 --RETLKRKVSGNR--CASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                  | :|  | | |   | |  ||| :    ||:| | :|| ||||:|||||| |||
Db          173 SGPPEDVKPPVLGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFK 230

Qy          175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRS 234
                |:|:    | |     ||:|| :|  || || :||   || :   : |         |||
Db          231 RTIRRDLTYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEE----  --RQRG 282

Qy          235 ADEQLHCACKAKRSGGHAPRVRELLLLDALSPEQL-VLTLLEAEPPHVLI      - 282
                |:    |  : :||                  :|||:: |  :||||   : :
Db          283 KDKD----GDGEGAGG-----  - APEEMPVDRILEAE--LAVEQKSDQGVEGP 322

Qy          283 -------SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
                      | |: | |      :: : |||:|  :: |||:|| |  | | ||| || : |
Db          323 GGTGGSGSSPNDPVT------NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 376

Qy          336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQH 394
                 |:|:     ||||    :: | | : |:       |:  ||| :|    |: |::::
Db          377 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDK 435

Qy          395 KEYLCVKAMILLNSSMYPLVTAFQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR 454
                |  |::|:|| | |     |        |:    :|     | :    |     ||   |
Db          436 TELGCLRAIILFNPDAKGL--------SNPSEVEVLREKVYASLETYCKQKYPEQQG--R 485

Qy          455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
                | ||: |  :|    | :|||    |      |:   |:||| |
Db          486 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 528


RESULT 34
US-10-098-184-2
; Sequence 2, Application US/10098184
; Publication No. US20030105333A1
; GENERAL INFORMATION:
;  APPLICANT: Pfahl, Magnus
;  APPLICANT:  Tachdjian, Catherine
;  APPLICANT:  Al-Shamma, Hussien A.
;  APPLICANT:  Fanjul, Andrea
;  APPLICANT:  Pleynet, David P.M.
;  APPLICANT:  Spruce, Lyle W.
;  APPLICANT:  Fine, Richard
;  APPLICANT:  Zapf, James W.
;  TITLE OF INVENTION: EXE ACTIVATING MOLECULES
;  FILE REFERENCE: 13099.001602
;  CURRENT APPLICATION NUMBER: US/10/098,184
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: 60/274,342
;  PRIOR FILING DATE: 2001-03-08
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 533
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20030105333A1e
;   OTHER INFORMATION: synthetic construct
US-10-098-184-2

  Query Match             16.3%;  Score 458.5;  DB 9;  Length 533;
  Best Local Similarity   40.2%;  Pred. No. 1.3e-41;
  Matches  140;  Conservative   63;  Mismatches  163;  Indels   97;  Gaps   17;

Qy           62 VTNLEGGPGRQTTSPNVILWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN- 120
                :::  | ||    :|     ||                  :: |  ||   :    :||
Db          136 ISSSMGSPGLPPPAP------PG--------  ---FSGPVSSPQINSTV---SLPGGG 172

Qy          121 ---RETLKRKVSGNR---CASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                   | :|  | | |   | |  ||| :    ||:| | :|| ||||:|||||| |||
Db          173 SGPPEDVKPPVLGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFK 230

Qy          175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRS 234
                |:|:    | |     ||:|| :|  || || :||   || :   : |         |||
Db          231 RTIRRDLTYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEE-----  --RQRG 282

Qy          235 ADEQLHCACKAKRSGGHAPRVRELLLLDALSPEQL-VLTLLEAEPPHVLI      - 282
                |:    |  : :||                  :|||:: |  :||||   : :
Db          283 KDKD----GDGEGAGG-------------APEEMPVDRILEAE--LAVEQKSDQGVEGP 322

Qy          283       SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
                      | |: | |      :: : |||:|  :: |||:|| |  | | ||| || : |
Db          323 GGTGGSGSSPNDPVT-----  NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 376

Qy          336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQH 394
```

```
                 |:|:     ||||    :: | | : |:      |:  ||| :|    |: |::::
Db        377 NELLIASFSHPSIDVRDGIIIATGLHVHRNSAHSA-GVGAIFDRVLIELVSKMRDMRMNF 435

Qy        395 KEYLGVKAMILLNSSMYPLVTATQDADSSRKIAHLLNAVTDALVWVIAKSGISSQQQSMR 454
              |  |::|:|| |     |        |:    :|    | :    |    ||   |
Db        436 TELGCLRAIILFNPDAKGL--------SNPSEVEVLREKVYASLETYCKQKYPEQQG--R 485

Qy        455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
              | ||: |  :|    | :|||   |     |:   |:||| |
Db        486 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 528


RESULT 35
US-09-853-386-92
; Sequence 92, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
;  APPLICANT: Murphy, Evelyn
;  APPLICANT:  Bresnihan, Barry
;  APPLICANT:  Conneely, Orla
;  APPLICANT:  Fitzgerald, Oliver
;  TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR
;  TITLE OF INVENTION:  Subfamily of Nuclear Transcription Factors
;  FILE REFERENCE: P01972US1
;  CURRENT APPLICATION NUMBER: US/09/853,386
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/203645
;  PRIOR FILING DATE: 2000-05-12
;  NUMBER OF SEQ ID NOS: 153
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 92
;   LENGTH: 601
;   TYPE: PRT
;   ORGANISM: HUMAN
US-09-853-386-92

  Query Match             16.0%;  Score 447.5;  DB 10;  Length 601;
  Best Local Similarity   28.9%;  Pred. No. 1.4e-30;
  Matches  158;  Conservative   63;  Mismatches  211;  Indels  115;  Gaps   20;

Qy          5 NSPSSLNSPSS----------YNCSQSIL--PLEHGSIYIPSSYVDSHHEYPAMTFYSPA 52
              :| ||  ||:|          | |    |  ||:       | |  |    |     ||:
Db         90 SSSSSATSPASASFKFEDFQVYGCYPGTLSGPIDETLSSSGSEYYGSPCSAP-----SPS 144

Qy         53 VMNYSIPSNVTNLEGGPGRQTTSPNV--LW----------------PT------PGHLSP 88
              |:  || ::  :|  |  : |    ||                    ||      |  ||
Db        145 TPNFQ-PSQLSPWDGSFGHFSPSQTYEGLWAWTEQLPKASSGPPPPPTFFSFSPPTGPSP 203

Qy         89 IVVHRQLSHIYAFPQKSPWCEARSLEHTLPV----------NRET---LKRKVSGNRCAS 135
              :   |  |:  |       | |  :::|              ||:|   |  |:  : |
Db        204 SLAQSSLK-LFPPPATHQLGEGES--YSMPAAFPGLAPTSPNRDTSGTLDAPVTSTKSRS 260

Qy        136 PVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDK 195
              :|    |    |||| | ||  |||| :|||||  ||||::|    ||| |   | :||
Db        261 GASGGSEGR----CAVCGDNASGHYGVRTCEGCKGFFKRTVQKSAKYICLANKQCPVDK 316

Qy        196 NRRFSQACRLRKCYEVGMVKGGSRFERGRMVKECRARQDECAVRMVEILGVEGKAIKGLE 255
              ||  || || :||  |||||           :||                |   | |
Db        317 RRRNRCQFCRFQKCLAVGMVK----------EVVR----------TDSLKGRRGRLPSK 365

Qy        256 RELLIDALSPPQLVTLLEAEPPHVLISKPSAPFTEASMMMSLT---------------- 299
              :   || ||  |: :|: |     | | ||    : |    |
Db        366 PKQPPDA-SPINLLTSLIRAH----LDSGPSTAKLDYSKFQELVLPRFGKEDAGDVQQFY 420

Qy        300 KLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPD 359
              |    |  :  ||:|||||:||   ||  |||| ::|: :: | :||    |||||
Db        421 DLTSGSLDVIRKWAEKIPGFIELCPGDQDLLLESAFLELFILRLAYRSKPGEGKLIFCSG 480

Qy        360 LVIDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQD 419
              |||   : :|  |  :  | :|| :     |  :    : |: |::          |:|
Db        481 LVL--HQLQCARFGDWIDNILAFSRSIHSLGVDVPAFACLSALV--------LITDRHG 530

Qy        420 ADSSRKIAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNM 479
              |:: | | :  |   :|     | |  |: || |  :|    :|::  :  :
Db        531 LQDPRRVEELQNRIASCLKEHMATVAGDPQPASC-LSRLLGKLPELRTLCTQGLQRIFCL 589

Qy        480 KCKNVVP 486
              | :::||
Db        590 KLEDLVP 596


RESULT 36
US-09-853-386-94
; Sequence 94, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
;  APPLICANT: Murphy, Evelyn
;  APPLICANT:  Bresnihan, Barry
;  APPLICANT:  Conneely, Orla
;  APPLICANT:  Fitzgerald, Oliver
;  TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR
;  TITLE OF INVENTION:  Subfamily of Nuclear Transcription Factors
;  FILE REFERENCE: P01972US1
;  CURRENT APPLICATION NUMBER: US/09/853,386
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/203645
;  PRIOR FILING DATE: 2000-05-12
;  NUMBER OF SEQ ID NOS: 153
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 94
;   LENGTH: 601
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-853-386-94

  Query Match             16.0%;  Score 447.5;  DB 10;  Length 601;
  Best Local Similarity   28.9%;  Pred. No. 1.4e-30;
  Matches  158;  Conservative   63;  Mismatches  211;  Indels  115;  Gaps   20;

Qy          5 NSPSSLNSPSS----------YNCSQSIL--PLEHGSIYIPSSYVDSHHEYPAMTFYSPA 52
              :| ||  ||:|          | |    |  ||:       | |  |    |     ||:
Db         90 SSSSSATSPASASFKFEDFQVYGCYPGTLSGPLDEILSSSGSEYYGSPCSAP-----SPS 144

Qy         53 VMNYSIPSNVTNLEGGPGRQTTSPNV--LW----------------PT------PGHLSP 88
              |:  || ::  :|  |  : |    ||                    ||      |  ||
Db        145 TPNFQ-PSQLSPWDGSFGHFSPSQTYEGLWAWTEQLPKASSGPPPPPTFFSFSPPTGPSP 203

Qy         89 IVVHRQLSHIYAFPQKSPWCEARSLEHTLPV----------NRET---LKRKVSGNRCAS 135
              :   |  |:  |       | |  :::|              ||:|   |  |:  : |
Db        204 SLAQSSLK-LFPPPATHQLGEGES--YSMPAAFPGLAPTSPNRDTSGTLDAPVTSTKSRS 260

Qy        136 PVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDK 195
              :|    |    |||| | ||  |||| :|||||  ||||::|    ||| |   | :||
Db        261 GASGGSEGR----CAVCGDNASGHYGVRTCEGCKGFFKRTVQKSAKYICLANKQCPVDK 316

Qy        196 NRRFSQACRLRKCYEVGMVKGGSRFERGRMVKECRARQDECAVRMVEILGVEGKAIKGLE 255
              ||  || || :||  |||||           :||                |   | |
Db        317 RRRNRCQFCRFQKCLAVGMVK----------EVVR----------TDSLKGRRGRLPSK 365

Qy        256 RELLIDALSPPQLVTLLEAEPPHVLISKPSAPFTEASMMMSLT---------------- 299
              :   || ||  |: :|: |     | | ||    : |    |
Db        366 PKQPPDA-SPINLLISLIRAH----LDSGPSTAKLDYSKFQELVLPRFGKEDAGDVQQFY 420

Qy        300 KLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPD 359
              |    |  :  ||:|||||:||   ||  |||| ::|: :: | :||    |||||
Db        421 DLTSGSLDVIRKWAEKIPGFIELCPGDQDLLLESAFLELFILRLAYRSKPGEGKLIFCSG 480

Qy        360 LVIDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQD 419
              |||   : :|  |  :  | :|| :     |  :    : |: |::          |:|
Db        481 LVL--HQLQCARFGDWIDNILAFSRSIHSLGVDVPAFACLSALV--------LITDRHG 530

Qy        420 ADSSRKIAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNM 479
              |:: | | :  |   :|     | |  |: || |  :|    :|::  :  :
Db        531 LQDPRRVEELQNRIASCLKEHMATVAGDPQPASC-LSRLLGKLPELRTLCTQGLQRIFCL 589

Qy        480 KCKNVVP 486
              | :::||
Db        590 KLEDLVP 596
```

```
                | :::||
Db          580 KLEDLVP 586


RESULT 47
US-09-922-226-3
; Sequence 3, Application: US/09922226
; Publication No. US20030077664A1
; GENERAL INFORMATION:
;  APPLICANT: Zhao, Yi
;  APPLICANT:  Thacher, Scott M.
;  APPLICANT:  Xiao, Jia-Hao
;  APPLICANT:  Kusari, Jyotirmoy
;  APPLICANT:  Chandraratna, Roshantha A.
;  TITLE OF INVENTION: Methods of Screening For Compounds That
;  TITLE OF INVENTION:  Modulate Hormone Receptor Activity
;  FILE REFERENCE: P-AR 4681
;  CURRENT APPLICATION NUMBER: US/09/922,226
;  CURRENT FILING DATE:  2002-01-09
;  PRIOR APPLICATION NUMBER: US 60/284,797
;  PRIOR FILING DATE: 2001-04-18
;  NUMBER OF SEQ ID NOS: 191
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 463
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-226-3

  Query Match             15.9%;  Score 446;  DB 9;  Length 463;
  Best Local Similarity   28.3%;  Pred. No. 1.3e-40;
  Matches  143;  Conservative   71;  Mismatches  215;  Indels   76;  Gaps   15;

Qy          7 PSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSTPSNVTNLE 66
              |:       :  | |: |    |   |   :|||  |      :|  :  |      |
Db         12 PAGYGGSPGHTGSTSMSPSAALSTGKP---MDSHPSYTDTPVSAPRTL--SAVGTPLNAL 66

Qy         67 GGPGRQTTSPNVLWPTPGHLS--PLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETL 124
              | | |  ||   : |  | |:  | :       :| | |                  ::  :
Db         67 GSPYRVITS--AMGPPSGALAAPPGI------NLVAPPS------- ------- SQLNV 102

Qy        125 RRKVSGNRCASPV-------------TGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKA 171
               || :   |:                | |||    | ||:| | :|| ||||:||||||
Db        103 VNSVSSSEDIKPLPGLPGTGNMNYPSTSPGS-LVKHICAICGDRSSGKHYGVYSCEGCKG 161

Qy        172 FFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGRLVRR 231
              ||||:|:    | |     | ||| :|  || || :||  :|| :   : ||      |
Db        162 FFKRTIRKDLIYTCRDNKDCLIDKRQRNRCQYCRYQKCLVMGMKREAVQEER-----QRS 216

Qy        232 QRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTE 291
              :  |: :  ||     | ::|:|   | |    |:      : :  : | |
Db        217 RERAESEAECA-----TSGHEDMPVERILEA---ELAVEPKTESYGDMNMENSTNDPVT  267

Qy        292 ASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHP 351
                ::  |||:|  :: |||:|| | :|:| ||| || : | |:|:     ||:
Db        268   --NICHAADKQLFTLVEWAKRIPHFSDLTLEDQVILLRAGWNELLIASFSHRSVSVQ 322

Qy        352 GKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSSM 410
              ::  |  | : |       |:  ||| :|    |: ::::   |  |::|:|: | |
Db        323 DGILLATGLHVHRSSAHSA-GVGSIFDRVLTELVSKMKDMQMDKSELGCLRAIVLFNPDA 381

Qy        411 YPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASN 470
              |   ::      |:  | | |                :|  | | ||: |  :|
Db        382 EGLSNPSEVFTLREKVYATLEAYTKQ---------KYPEQPGRFAKLLLRLPALRSIGL 431

Qy        471 KGMEHLLNMKCKNVVPVYDLLLEML 495
              | :|||   |     |:   |:|||
Db        432 KCLEHLFFFKLIGDTPIDTFLMEML 456


RESULT 48
US-10-098-184-3
; Sequence 3, Application US/10098184
; Publication No. US20030105333A1
; GENERAL INFORMATION:
;  APPLICANT: Pfahl, Magnus
;  APPLICANT:  Tachdjian, Catherine
;  APPLICANT:  Al-Shamma, Hussien A.
;  APPLICANT:  Fanjul, Andrea
;  APPLICANT:  Pleynet, David P.M.
;  APPLICANT:  Spruce, Lyle W.
;  APPLICANT:  Fine, Richard
;  APPLICANT:  Zapf, James W.
;  TITLE OF INVENTION: RXR ACTIVATING MOLECULES
;  FILE REFERENCE: 13099.0016U2
;  CURRENT APPLICATION NUMBER: US/10/098,184
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: 60/274,342
;  PRIOR FILING DATE: 2001-03-08
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 463
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence; No. US20030105333A1e
;   OTHER INFORMATION: synthetic construct
US-10-098-184-3

  Query Match             15.9%;  Score 446;  DB 9;  Length 463;
  Best Local Similarity   28.3%;  Pred. No. 1.3e-40;
  Matches  143;  Conservative   71;  Mismatches  215;  Indels   76;  Gaps   15;

Qy          7 PSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSTPSNVTNLE 66
              |:       :  | |: |    |   |   :|||  |      :|  :  |      |
Db         12 PAGYGGSPGHTGSTSMSPSAALSTGKP---MDSHPSYTDTPVSAPRTL--SAVGTPLNAL 66

Qy         67 GGPGRQTTSPNVLWPTPGHLS--PLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETL 124
              | | |  ||   : |  | |:  | :       :| | |                  ::  :
Db         67 GSPYRVITS--AMGPPSGALAAPPGI------NLVAPPS------- ------- SQLNV 102

Qy        125 RRKVSGNRCASPV-------------TGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKA 171
               || :   |:                | |||    | ||:| | :|| ||||:||||||
Db        103 VNSVSSSEDIKPLPGLPGTGNMNYPSTSPGS-LVKHICAICGDRSSGKHYGVYSCEGCKG 161

Qy        172 FFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGRLVRR 231
              ||||:|:    | |     | ||| :|  || || :||  :|| :   : ||      |
Db        162 FFKRTIRKDLIYTCRDNKDCLIDKRQRNRCQYCRYQKCLVMGMKREAVQEER-----QRS 216

Qy        232 QRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTE 291
              :  |: :  ||     | ::|:|   | |    |:      : :  : | |
Db        217 RERAESEAECA-----TSGHEDMPVERILEA---ELAVEPKTESYGDMNMENSTNDPVT  267

Qy        292 ASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHP 351
                ::  |||:|  :: |||:|| | :|:| ||| || : | |:|:     ||:
Db        268   --NICHAADKQLFTLVEWAKRIPHFSDLTLEDQVILLRAGWNELLIASFSHRSVSVQ 322

Qy        352 GKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSSM 410
              ::  |  | : |       |:  ||| :|    |: ::::   |  |::|:|: | |
Db        323 DGILLATGLHVHRSSAHSA-GVGSIFDRVLTELVSKMKDMQMDKSELGCLRAIVLFNPDA 381

Qy        411 YPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASN 470
              |   ::      |:  | | |                :|  | | ||: |  :|
Db        382 EGLSNPSEVFTLREKVYATLEAYTKQ---------KYPEQPGRFAKLLLRLPALRSIGL 431

Qy        471 KGMEHLLNMKCKNVVPVYDLLLEML 495
              | :|||   |     |:   |:|||
Db        432 KCLEHLFFFKLIGDTPIDTFLMEML 456


RESULT 49
US-09-965-703-63
; Sequence 63, Application US/09965703
; Patent No. US20020119521A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Rohm and Haas Company
;  APPLICANT:  Palli, Subba Reddy
;  APPLICANT:  Kapitskaya, Marianna Zinovjevna
;  APPLICANT:  Cress, Dean Ervin
;  TITLE OF INVENTION: No. US20020119521A1o1 Polypeptide Based Insecticide Tar
;  FILE REFERENCE: A01020B
;  CURRENT APPLICATION NUMBER: US/09/965,703
;  CURRENT FILING DATE:  2001-09-26
;  PRIOR APPLICATION NUMBER: 60/191,355
;  PRIOR FILING DATE: 2000-03-22
;  PRIOR APPLICATION NUMBER: 60/269,799
;  PRIOR FILING DATE: 2001-02-20
;  PRIOR APPLICATION NUMBER: PCT/US01/09050
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63
;   LENGTH: 467
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: No. US20020119521A1o1 Sequence
US-09-965-703-63

  Query Match             15.8%;  Score 442.5;  DB 10;  Length 467;
  Best Local Similarity   28.3%;  Pred. No. 2.7e-30;
  Matches  150;  Conservative   78;  Mismatches  209;  Indels  101,  Gaps   18;

Qy          1 MDIKN----------SPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVD------------ 38
              || |:           : ||||||:           ||: :|| :
Db          1 MDTKHFLPLDFSTQVNSSSLNSPTG----------RGSMAVPSLHPSLGPGIGSPLGSP 49

Qy         39 SHHFYPAMTFYSPAVMNYSIPSNVTNLPGGPGPQT--TSPNVLWPT--PGHLSPLVVHRQ 94
                 |  | || :     | :| :   ||    .  |:| . . |  |   ||:
Db         50 GQLHSPISTLSSP-INGMGPPFSVISSPMGPHSMSVPTTPTLGFGTGSPQLNSPM----- 103

Qy         95 LSHLYAEPQKSPWCEARSLEHTLPVNRFTLK--RKVSGNRCASPVTGPGSKRDAHFCAVC 152
                      :|      ::  | :|   ||     |||  :           | ||:|
Db        104 ----------NPVSSTEDIKPPLGLN-GVLKVPAHPSGNMASF---------TKHICAIC 143

Qy        153 SDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEV 212
               | :|| ||||:|||||| ||||:::    | |     | ||| :|  || || :||  :
Db        144 GDRSSGKHYGVYSCEGCKGFFKRTVRKDLTYTCRDNKDCLIDKRQRNRCQYCRYQKCLAM 203

Qy        213 GMVKCGSPRPRCGYPLVPPQRSADFQLHCAGKAFPSGCHAPPVPELLLDALSPEQLVLTL 272
              || :   : |          ||| |   :       :    | | ::|   |: |     |
Db        204 GMKREAVQEE--------RQRGKDRNENEVESTSSANEDMP-VEKILEAELAVEPKTETY 254

Qy        273 LEA----EPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQV 328
              :||     |     | |: | |      :: : |||:|  :: |||:|| | || | |||
Db        255 VEANMGLNP-----SSPNDPVT------NICQAADKQLFTLVEWAKRIPHFSELPLDDQV 303

Qy        329 RLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRF 387
               || : | |:|:     |||     :: | | : |:      |:  ||| :|    |:
Db        304 ILLRAGWNELLIASFSHRSIAVKDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKM 362

Qy        388 RELKIQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGIS 447
              |:::   |  |::|::| |     |   :     |:   | |
Db        363 RDMQMDKTELGCLRAIVLFNPDSKGLSNPAEVEALREKVYASLEAYCKH----------K 412

Qy        448 SQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
               :|  | | ||: | :|    | :|||   |     |:   |:||| |
Db        413 YPEQPGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 462


RESULT 40
US-10-211-239-2
; Sequence 2, Application US/10211239
; Publication No. US20030103965A1
; GENERAL INFORMATION:
;  APPLICANT: Jung, Birgit
;  APPLICANT:  Kraut, No  US20030103965Albert
;  APPLICANT:  Mueller, Stefan
;  TITLE OF INVENTION: Method for Identifying Substances Which Positively
;  TITLE OF INVENTION:  Influence Inflammatory Conditions
;  FILE REFERENCE: 0652.2340001
;  CURRENT APPLICATION NUMBER: US/10/211,239
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/315,775
;  PRIOR FILING DATE: 2001-08-30
;  PRIOR APPLICATION NUMBER: EP 01119008.0
;  PRIOR FILING DATE: 2001-08-06
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 598
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-211-239-2

  Query Match             15.8%;  Score 442.5;  DB 9;  Length 598;
  Best Local Similarity   28.8%;  Pred. No. 3.8e-30;
  Matches  152;  Conservative   58;  Mismatches  202;  Indels  115;  Gaps   17;

Qy          2 DIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTF------------ 48
              |  || |  |||: :     |    ||              |  |: |:
Db        130 DYYGSPCSAPSPSTPSFQPPQLSPWDGSF---------GHFSPSQIYEGLRAWTEQLPKA 180

Qy         49   - YSPAVMNYSIPSNVTNLEGGPGRQ-TTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQ 103
                  ||  ..| |:       ||      ||  |:|:           | :|   |
Db        181 SGPPQPPAFFSFSPPT-------GPSPSLAQSPLKLFPS----------QATHQLGEG- 221

Qy        104 KSPWCEARSLEHTLPVNRFTLKRKVSGNRCASPVT-------GPGSKRDAHFCAVCSDYA 156
                    |: |:    |    |           :|||        ||       |||| | |
Db        222 -----ESYSMPTAFPGLAFTSPHLEGSGILDIFVTSTKARSGAPGGSEGR--CAVCGDNA 274

Qy        157 SGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRPKSCQACRLRKCYEVGMVK 216
              |  |||| :|||||  ||||:::| :  ||| |   | :|| ||  || || :||  |||||
Db        275 SCQHYGVRTCEGCKGFFKRIVQKNAKYICLANKDCPVDKRRRNRCQFCRFQKCLAVGMVK 334

Qy        217 CGSPRPRCGYPLVPPQRSADFQLHCAGKAFPSGCHAPPVPELLLDALSPEQLVLTLIPAF 276
                        :||           |  |  |   :   || ||  |: :|: |
Db        335 ----------EVVR-----------TDSLKGRRGRLPSKPKQPPDA-SPANLLTSLVRAH 372

Qy        277 PPHVLISRPSAPFTEASMMMSLT----------------KLADKELVHMISWAKKIPGFV 320
                | | ||    : |    |                  |    |  :  ||:||||
Db        373 ----LDSGPSTAKLDYSKFQELVLPHFGKEDAGDVQQFYDLLSGSLEVIRKWAEKIPGFA 428

Qy        321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
              |||  ||  |||| ::|: :: | :||    |||||   ||| | :  | |  :  | :
Db        429 ELSPADQDLLLESAFLELFILGLAYRSKPGEGKLIFCSGLVLHRLQ--CARGFGDWIDSI 486

Qy        381 LATTSRFRELKIQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWV 440
              || :     | :     : |::       :|         |::  | | :    |
Db        487 LAFSRSLHSLLVDVPAFACLSALV--------LITDRHGLQEPRRVEELQNRIASCLKEH 538

Qy        441 VIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVP 486
              | | :|    | :  |: || |  :|    :||: :  :| ::::||
Db        539 VAAVAG--EPQPASCLSRLLGKLPELRTLCTQGLQRIFYLKLEDLVP 583


Search completed: June 16, 2003, 13:41:54
Job time : 59.7301 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:52:29 ; Search time 29.4238 Seconds
(without alignments)
1731.636 Million cell updates/sec

Title: US-08-826-361A-25
Perfect score: 2805
Sequence: 1 MDIKNSPSSINSPSSYNCSQ ... PCSPAEDSKSKECSQNPQSQ 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2805	100.0	530	2	JC5939	estrogen receptor
2	2522	89.9	477	2	S71400	estrogen receptor
3	2284	81.4	503	2	JW0046	estrogen receptor
4	1230.5	43.9	589	1	QRCHE	estrogen receptor
5	1224.5	43.7	595	1	QRHUE	estrogen receptor
6	1219.5	43.5	595	2	I47140	estradiol receptor
7	1218.5	43.4	599	1	QRMSE	estrogen receptor
8	1216.5	43.4	600	1	QRRTE	estrogen receptor
9	1214	43.3	586	1	QRXIE	estrogen receptor
10	1199	42.7	620	2	T10423	estrogen receptor
11	1174.5	41.9	701	2	S64737	80K estrogen recep
12	1134.5	40.4	535	2	S58224	oestrogen receptor
13	1120	39.9	574	2	A47197	estrogen receptor
14	631.5	22.5	433	2	B29345	steroid hormone re
15	619.5	22.1	433	2	S58087	estrogen receptor
16	597.5	21.3	521	2	A29345	steroid hormone re
17	471	16.8	467	2	A43781	retinoid-X-recepto
18	469.5	16.7	923	2	A39596	progesterone recep
19	465	16.6	923	2	I53280	progesterone recep
20	459	16.4	446	2	A34418	H-2 region II bind
21	459	16.4	451	2	A41651	retinoic acid rece
22	458.5	16.3	533	2	S37781	retinoid X recepto
23	457.5	16.3	758	2	S60586	glucocorticoid rec
24	456	16.3	448	2	D41727	retinoid X recepto
25	456	16.3	520	2	I84718	RXR-beta1 isoform
26	455.5	16.2	410	2	S26669	retinoic acid rece
27	453	16.1	848	2	JQ0194	androgen receptor
28	453	16.1	930	2	A25923	progesterone recep
29	452	16.1	463	2	S26670	retinoic acid rece
30	452	16.1	463	2	B41727	retinoid-X recepto
31	451	16.1	933	1	QRHUP	progesterone recep
32	447.5	16.0	601	1	QRMSN1	probable hormone r
33	447.5	16.0	984	2	A29513	mineralocorticoid
34	446	15.9	441	2	I50515	retinoid X recepto
35	442.5	15.8	467	2	A47278	retinoid X recepto
36	442.5	15.8	467	2	S26668	retinoic acid rece
37	442.5	15.8	598	2	A37251	probable nuclear h
38	442.5	15.8	786	2	A35466	progesterone recep
39	441.5	15.7	776	1	S44047	glucocorticoid rec
40	441.5	15.7	981	2	A41401	mineralocorticoid
41	441	15.7	470	2	D41977	retinoid receptor
42	441	15.7	563	2	JQ0623	nerve growth facto
43	438.5	15.6	777	1	QRBUGA	glucocorticoid rec
44	438	15.6	462	1	S09592	retinoid X recepto
45	438	15.6	899	2	A35895	androgen receptor
46	438	15.6	902	2	B40494	androgen receptor
47	437.5	15.6	919	2	A39248	androgen receptor
48	435	15.5	910	2	A34721	androgen receptor
49	434.5	15.5	911	2	B34721	androgen receptor
50	432.5	15.4	771	2	A54273	glucocorticoid rec
51	431.5	15.4	742	1	QRHUGB	glucocorticoid rec
52	427.5	15.2	452	2	S47633	RXR protein - Atri
53	427.5	15.2	577	2	S33763	hormone receptor N
54	425.5	15.2	462	2	S44490	RXR type hormone r
55	424.5	15.1	438	2	I50517	retinoid X recepto
56	423.5	15.1	783	1	A25691	glucocorticoid rec
57	419	14.9	379	2	I50514	retinoid X recepto
58	417.5	14.9	795	1	QRRTG	glucocorticoid rec
59	416	14.8	488	2	C41977	retinoid receptor
60	413.5	14.7	422	2	I50516	retinoid X recepto
61	411	14.7	80	2	PW0044	estrogen receptor
62	410	14.6	385	2	I53158	orphan nuclear rec
63	407	14.5	385	2	S48074	nuclear receptor T
64	404.5	14.4	344	2	I51330	androgen receptor
65	403.5	14.4	508	2	S11513	usp protein - frui
66	403.5	14.4	508	2	T13737	steroid hormone re
67	398.5	14.2	598	2	I84692	nuclear orphan rec
68	397.5	14.2	495	2	A57053	germ cell nuclear
69	397	14.2	598	2	A57040	T-cell nuclear rec
70	396	14.1	459	2	A41977	retinoic acid rece
71	394.5	14.1	598	2	A46225	nuclear orphan rec
72	384	13.7	560	2	S27874	steroid hormone re
73	383	13.7	448	2	B56558	retinoic acid rece
74	382.5	13.6	133	2	S26595	estrogen receptor
75	380.5	13.6	464	2	A56558	retinoic acid rece
76	377.5	13.5	628	2	JC2493	serum derived orp
77	376.5	13.4	808	1	S33708	nuclear steroid/th
78	376	13.4	458	2	A34714	retinoic acid rece
79	376	13.4	462	2	S05050	retinoic acid nucl
80	373.5	13.3	447	2	B34714	retinoic acid rece
81	372.5	13.3	443	1	C35991	retinoic acid rece
82	372	13.3	455	2	S13512	retinoic acid rece
83	372	13.3	455	2	A36471	transcription fact
84	372	13.3	462	1	A56120	steroid hormone re
85	370.5	13.2	454	1	A33903	retinoic acid rece
86	370	13.2	501	1	A56543	Ftz-F1 related orp
87	369.5	13.2	414	2	A37133	apolipoprotein A-I
88	369.5	13.2	414	2	I48975	apolipoprotein A-I
89	369.5	13.2	461	1	A46077	steroid hormone re
90	369.5	13.2	774	2	JC6095	hepatocyte nuclear
91	369	13.2	465	2	S52074	hepatocyte nuclear
92	369	13.2	465	2	S23502	hepatocyte nuclear
93	368.5	13.1	403	2	S35384	steroid receptor p
94	367.5	13.1	444	2	I51256	retinoic acid rece
95	367.5	13.1	453	2	I50674	retinoic acid rece
96	367.5	13.1	454	2	S06124	retinoic acid rece
97	367.5	13.1	625	2	S71930	neuron-derived rec
98	367	13.1	462	1	A29491	retinoic acid rece
99	367	13.1	462	2	A40716	steroid hormone re
100	366	13.0	381	2	A46458	steroid/thyroid ho
101	366	13.0	458	2	S06123	retinoic acid rece
102	365.5	13.0	504	2	JC6096	hepatocyte nuclear

103	465	13.0	410	2	I50689	orphan receptor CO
104	465	13.0	448	2	S02827	retinoic acid rece
105	464	13.0	561	2	S58238	DHR38 protein - fr
106	463.5	13.0	452	2	S78481	retinoic acid rece
107	463	12.9	476	2	B41977	retinoic acid rece
108	462.5	12.9	448	2	S05051	retinoic acid rece
109	462	12.9	408	2	JC4938	hepatocyte nuclear
110	462	12.9	465	2	JC4937	hepatocyte nuclear
111	460.5	12.9	461	2	JC4972	steroidogenic fact
112	460	12.8	423	2	S02710	transcription fact
113	460	12.8	442	2	A38592	retinoic acid rece
114	460	12.8	447	2	I38975	nuclear orphan rec
115	459.5	12.8	411	2	S35333	steroid receptor p
116	459	12.8	422	2	I48405	gene FCUP TF1 prot
117	458	12.8	543	2	A32693	steroid receptor p
118	456.5	12.7	448	2	A43786	retinoic acid rece
119	456.5	12.7	461	2	JC4014	steroid hormone-nu
120	455.5	12.7	419	2	A57000	orphan receptor CO

ALIGNMENTS

RESULT 1
JC5939
estrogen receptor beta - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C;Accession: JC5939
R;Ogawa, S.; Inoue, S.; Watanabe, T.; Hiroi, H., Orimo, A.; Hosoi, T., Ouchi, Y.; Muram
Biochem. Biophys. Res. Commun. 243, 122-126, 1998
A;Title: The complete primary structure of human estrogen receptor beta (hERbeta) and it
A;Reference number: JC5939; MUID:98139878; PMID:9473491
A;Accession: JC5939
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-530 <OGA>
A;Cross-references: DDBJ:AB006590; NID:g2911151; PIDN:BAA24953.1; PID:g2911152
C;Superfamily: estrogen receptor; erbA transforming protein homology
F;147-408/Domain: erbA transforming protein homology <ERB>

Query Match 100.0%; Score 2805; DB 2; Length 530;
Best Local Similarity 100.0%; Pred. No. 4e-211;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRRQRSADEQLH 240

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
```

RESULT 2
S71400
estrogen receptor beta - human
C;Species: Homo sapiens (man)
C;Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Jun-2000
C;Accession: S71400
R;Mosselman, S.; Polman, J.; Dijkema, R.
FEBS Lett. 392, 49-53, 1996
A;Title: ER-beta: identification and characterization of a novel human estrogen recep
A;Reference number: S71400; MUID:96354875; PMID:8769313
A;Accession: S71400
A;Molecule type: mRNA
A;Residues: 1-477 <MOS>
A;Cross-references: EMBL:X99101; NID:g1518262; PIDN:CAA67555.1; PID:g1518263
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; phosphoprotein; steroid binding; steroid hormone re
F;94-355/Domain: erbA transforming protein homology <ERBA>
F;96-116/Region: zinc finger CCCC motif
F;132-156/Region: zinc finger CCCC motif
F;167-182/Region: nuclear location signal
F;92/Binding site: phosphate (Ser) (covalent) (by MAP kinase) #status predicted
F;96,99,113,116/Binding site: zinc (Cys) #status predicted
F;132,138,148,151/Binding site: zinc (Cys) #status predicted

Query Match 89.9%; Score 2522; DB 2; Length 477;
Best Local Similarity 100.0%; Pred. No. 3.4e-189;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         54 MNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 113
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSL 60

Qy        114 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 173
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFF 120

Qy        174 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRRQR 233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRRQR 180

Qy        234 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEAS 240

Qy        294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK 300

Qy        354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 360

Qy        414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 420

Qy        474 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 477
```

RESULT 3
JW0046
estrogen receptor beta2 - rat

```
N;Alternate names: ERbeta2
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C;Accession: JW0046
R;Maruyama, K.; Endoh, H.; Sasaki-Iwaoka, H.; Kanou, H.; Shimaya, E.; Hashimoto, S.; Kat
Biochem. Biophys. Res. Commun. 246, 142-147, 1998
A;Title: A novel isoform of rat estrogen receptor beta with 18 amino acid insertion in t
A;Reference number: JW0046; MUID:98262932; PMID:9600083
A;Accession: JW0046
A;Molecule type: mRNA
A;Residues: 1-503 <MAR>
A;Cross-references: DDBJ:AB012721
C;Comment: This protein functions as a negative regulator of estrogen action.
C;Superfamily: estrogen receptor; erbA transforming protein homology
F;102-381/Domain: erbA transforming protein homology <ERBA>

  Query Match             81.4%;  Score 2284;  DB 2;  Length 503;
  Best Local Similarity   85.9%;  Pred. No. 1.4e-170;
  Matches  432;  Conservative   23;  Mismatches   30;  Indels   18;  Gaps    1;

Qy          46 MTFYSPAVMNYSIPSNVTNIFGGPGPQTTSPNVLWPTPCHLSPLVVHPQLSHLYAEPQKS 105
               ||||||||||||:| : :||:|||  ||:|||||||| ||||||  | | | ||||||||
Db           1 MTFYSPAVMNYSVPGSTSNIDGGPVRQSTSPNVLWPTSGHLSPLATHCQSSLLYAEPQKS 60

Qy         106 PWCPARSLEHTLPVNPETLKPKVSGNPCASPVTGPGSKPDAHPCAVCSDYASGYHYGVWS 165
               ||||||||||||||||||||||:||: |||||| | :|||||||||||||||||||||||
Db          61 PWCEARSLEHTLPVNPETLKPKLSGSSCASPVTSPNAKPDAHPCAVCSDYASGYHYGVWS 120

Qy         166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 225
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 180

Qy         226 YRLVRRQRSADEQLHCAGKAKRSGGHAPRVREILLLDALSPEQLVLTLLEAEPPHVLISRP 285
               ||:||||||: ||:|| ||||:|||||||:|||| ||||||||||||||||:||:|||
Db         181 YRIVRRQRSSSPQVHCLSKAKRNGGHAPPVKELLLSTISPEQLVLTLLEAEPPNVLVSPP 240

Qy         286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
               | ||||||||||||||||||| |||||||||||| ||||||||||||||||:|||
Db         241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy         346 RSIDHPGKLIFAPDLVLD------------------RDEGKCVEGILEIFDMLLATTSRF 387
               ||||||||||||||||                  ||||||||||||||||||||||
Db         301 RSIDHPGKLIFAPDLVLDRSSEDPHWHVAQMKSAAPRDEGKCVEGILEIFDMLLATTSRF 360

Qy         388 RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGIS 447
               |||||||||||||||||||||||| :| |:|:|||| ||||||||||||||||||||||
Db         361 RELKIQHKEYLCVKAMILLNSSMYPLASANQFAESSRKLTHLLNAVTDALVWVIAKSGIS 420

Qy         448 SQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSI 507
               |||||:|||||||||||||| ||||||||:|||||||||||||||||||| ||| ||||
Db         421 SQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEMLNAHILRGYKSSI 480

Qy         508 TGSECSPAEDSKSKEGSQNPQSQ 530
               :|||||  ||||:|| ||| |||
Db         481 SGSECSSTEDSKNKESSQNLQSQ 503


RESULT 4
ORCHE
estrogen receptor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 06-Mar-1992 #sequence_revision 14-Jul-1994 #text_change 22-Jun-1999
C;Accession: A40914; S07192
R;Maxwell, B.L.; McDonnell, D.P.; Conneely, O.M.; Schulz, T.Z.; Greene, G.L.; O'Malley,
Mol. Endocrinol. 1, 25-35, 1987
A;Title: Structural organization and regulation of the chicken estrogen receptor.
A;Reference number: A40914; MUID:88318621; PMID:2901032
A;Accession: A40914
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-589 <MAX>
R;Krust, A.; Green, S.; Argos, P.; Kumar, V.; Walter, P.; Bornert, J.M.; Chambon, P.
EMBO J. 5, 891-897, 1986
A;Title: The chicken oestrogen receptor sequence: homology with v-erbA and the human
A;Reference number: S07192; MUID:86247578; PMID:3755102
A;Accession: S07192
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-256,'E',258-589 <KRU>
A;Cross-references: EMBL:X03805; NID:g63378; PIDN:CAA27433.1; PID:g63380
C;Comment: The steroid hormones and their receptors are involved in the regulation of
C;Comment: In the absence of ligand, steroid hormone receptors are thought to be weak
omplex appears to recognize discrete DNA sequences upstream of transcriptional start
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; phosphoprotein; steroid binding; steroid hormone re
F;1-173/Domain: amino terminal <NH2>
F;174-255/Domain: DNA binding #status predicted <DNA>
F;177-450/Domain: erbA transforming protein homology <ERBA>
F;177-200/Region: zinc finger CCCC motif
F;213-235/Region: zinc finger CCCC motif
F;250-265/Region: nuclear location signal
F;294-546/Domain: steroid binding #status predicted <SIB>
F;179,182,196,199/Binding site: zinc (Cys) #status predicted
F;215,221,231,234/Binding site: zinc (Cys) #status predicted
F;230,299/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             43.9%;  Score 1230.5;  DB 1;  Length 589;
  Best Local Similarity   47.3%;  Pred. No. 3.6e-88;
  Matches  259;  Conservative   93;  Mismatches  126;  Indels   70;  Gaps   12;

Qy          12 SPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYS--IPSNVTNLEGGP 69
               :|:| :  | |   |   :| | |      | :  || : ::|  :|  : | :|
Db          77 APTSESPGSSSLAGPHSLNNVPPSPVVFLQTAPQL---SPFIHHHSQQVPYYLENEQGSF 133

Qy          70 GPQTTSPNVLWPTPCHLSPIVVHPQLSHLYAEPQKSPWCPARSLEHTLPVNPETLKPKVS 129
               | :  :|     |   |  :|                                | ::: :|
Db         134 GMREAAP------PAFYRPSSDNR-------------------------RHSIKRKMS 160

Qy         130 GNPCASPVTGPGSKPDAHPCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATN 189
                  ::   :| :  :||||:||||||||||||||||||||||||||||:|||||
Db         161 STNEKGSLSMESTK-ETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATN 219

Qy         190 QCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSADEQIHCAGKAKRSG 249
               |||||||||||||||||||||||:| | |::| | |::::|  :||    |:|  :
Db         220 QCTIDKNPPKSCQACPLRKCYEVGMMKGGIRKDRRGGRMMKQKPQPEEQQSRNGFASSTE 279

Qy         250 GHAPRV--REIIID---------ALSPEQLVLTLLEAEPPHVLISRPSAPFTFASMMMS 297
                 ||  :    |::           :|: ||:|  ||||||| |     |: || ||||
Db         280 LPAPTLWTSPLVVKHNKENSPALSLTAEQMVSALLEAEPPIVYSEYDPNRPFNEASMMTL 339

Qy         298 LTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFA 357
               || |||:|||||:|||::||||:|:| ||| |||  |:|:||:|||:::|||||:||
Db         340 LTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFA 399

Qy         358 PDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT 417
               |:|:|||::|||||:|||||||| :||| : || :|:|:|:|:|||| :|  :::|
Db         400 PNLITDRNQGKCVEGMVEIFDMLLATAARFRMMNLQGEEFVCLKSIILLNSGVYTFLSST 459

Qy         418 QDADSSPKLAH-LLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHL 476
                 :   |   | :|: :|| |: ::|||||:| |||  ||| |||::|||:|| ||||||||
Db         460 LKSLEEPDYIHPVLDKITDTLIHLMAKSGLSLQQQHRRLAQLLLILSHIRHMSNKGMEHL 519

Qy         477 LNMKCKNVVPVYDLLLEMLNAHVLR--GCKSSITGSE-------CSPAE---------DS 518
               |||||||:|||||||:|| |     :|:   |       :||          :|
Db         520 YNMKCKNVVPLYDLLLEMLDAHRLHAPAAKSAAEMEEPNRNQLITAPASSHSLQSFYINS 579

Qy         519 KSKEGSQN 526
               | :|  ||
Db         580 KEEESMQN 587


RESULT 5
ORHUE
estrogen receptor 1 - human
N;Alternate names: ER1; estrogen receptor alpha
C;Species: Homo sapiens (man)
```

```
C;Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 22-Jun-1999
C;Accession: A94284; A93476; A43021; S27143; S34000; A41925; B41925; A03244, 'A1925', 341
R;Greene, G.L.; Gilna, P.; Waterfield, M.; Baker, A.; Hort, Y.; Shine, J.
Science 231, 1150-1154, 1986
A;Title: Sequence and expression of human estrogen receptor complementary DNA.
A;Reference number: A94284; MUID:86122927; PMID:3753802
A;Accession: A94284
A;Molecule type: mRNA
A;Residues: 1-595 <GRI>
A;Cross-references: GB:M12674; NID:g182192; PIDN:AAA52399.1; PID:g182193
R;Green, S.; Walter, P.; Kumar, V.; Krust, A.; Bornert, J.M.; Argos, P.; Chambon, P.
Nature 320, 134-139, 1986
A;Title: Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A.
A;Reference number: A93476; MUID:86146892; PMID:3754034
A;Accession: A93476
A;Molecule type: mRNA
A;Residues: 1-595 <GR2>
A;Cross-references: GB:X03635; NID:g31233; PIDN:CAA27284.1; PID:g31234
R;Ponglikitmongkol, M.; Green, S.; Chambon, P.
EMBO J. 7, 3385-3388, 1988
A;Title: Genomic organization of the human oestrogen receptor gene.
A;Reference number: A43021; MUID:89091079; PMID:3145193
A;Accession: A43021
A;Molecule type: DNA
A;Residues: 143-161;205-225;244-264;356-374,'G',376;402-422;447-460,'P',462-462;508-528
R;Kenveney, M.; King, J.; Gannon, F.
DNA Seq. 2, 347-358, 1992
A;Title: Sequence analysis of the 5' flanking region of the human estrogen receptor gene
A;Reference number: S27140; MUID:93075998; PMID:1476547
A;Accession: S27143
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-115 <KEA>
A;Cross-references: EMBL:X62462; NID:g31201; PIDN:CAA44322.1; PID:g31205
R;Pfeffer, U.; Fecarotta, E.; Castagnetta, L.; Vidali, G.
Cancer Res. 53, 741-744, 1993
A;Title: Estrogen receptor variant messenger RNA lacking exon 4 in estrogen-responsive h
A;Reference number: S34000; MUID:93153765; PMID:7916651
A;Accession: S34000
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 216-254,367-399,'G',401-434 <PFE>
A;Cross-references: EMBL:X73067; NID:g579865; PIDN:CAA51528.1; PID:g939886
A;Note: the authors translated the codon GGG for residue 400 as Val
R;Dotzlaw, H.; Alkhalaf, M.; Murphy, L.C.
Mol. Endocrinol. 6, 773-785, 1992
A;Title: Characterization of estrogen receptor variant mRNAs from human breast cancers.
A;Reference number: A41925; MUID:92293154; PMID:1603086
A;Accession: A41925
A;Molecule type: mRNA
A;Residues: 1-214,'ELPTLC' <DOT>
A;Cross-references: GB:M69297; NID:g182218; PIDN:AAA58462.1; PID:g182219
A;Experimental source: clone 4; breast cancer
A;Note: sequence has been revised after extraction from NCBI backbone
A;Note: the complete sequence of neither the nucleotide nor the protein is shown in this
A;Note: sequence extracted from NCBI backbone (NCBIN:106580)
A;Accession: B41925
A;Status: significant sequence differences
A;Molecule type: mRNA
A;Cross-references: GB:M69296
A;Experimental source: clone 24; breast cancer
A;Note: sequence extracted from NCBI backbone (NCBIN:106597)
C;Comment: The steroid hormones and their receptors are involved in the regulation of eu
C;Comment: In the absence of ligand, steroid hormone receptors are thought to be weakly
omplex appears to recognize discrete DNA sequences upstream of transcriptional start sit
C;Genetics:
A;Gene: GDB:ESR1; ESR
A;Cross-references: GDB:119120, OMIM:133430
A;Map position: 6q25.1-6q25.1
A;Introns: 151/2; 215/3; 254/1; 309/1; 417/2; 477/1; 580/2
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; phosphoprotein; steroid hormone receptor; transcriptio
F;1-120/Domain: amino-terminal <NB2>
F;121-299/Domain: DNA binding #status predicted <DNA>
F;183-456/Domain: erbA transforming protein homology <ERA>
F;185-205/Region: zinc finger C4CC motif
F;221-245/Region: zinc finger C4CC motif
F;400-595/Domain: steroid binding #status predicted <STB>
F;236,305/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             43.7%;  Score 1224.5;  DB 1;  Length 595;
  Best Local Similarity   48.3%;  Pred. No. 1.1e-87;
  Matches  264;  Conservative   88;  Mismatches  129;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   |  :: |  |  | : |  | :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                 : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EARSLEHTLPVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSD 154
               : | :   ||       |    :  |: |    :      |     | ::  :||||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGSRREPCGYRLVRRQRSADEQLHCAGKAFEPSGCHAPRVR------------EL 258
              :| | |::| | |::: :|  |:       |: :  | |  :|
Db        251 MKGCIRKDRRGGRMLKHKRQRDD-----GEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :| |   |: ||:|||||  || |||:||||||:|||::
Db        306 LALSLTADQMVSALLDAEPP  ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:|||:|||::|| ||:|||:|:|||::|||||::||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||||:|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: :::|:|::: |||  ||| ||:::|||:|| |||||||| :|||||||||:|||||||
Db        485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy        496 NAHVL 500
              :||  |
Db        545 DAHRL 549


RESULT 6
I47140
estradiol receptor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Aug-1999
C;Accession: I47140; S66250; S32402
R;Bokenkamp, D.; Junqblut, P.W.; Thole, H.H.
Mol. Cell. Endocrinol. 104, 163-172, 1994
A;Title: The C-terminal half of the porcine estradiol receptor contains no post trans
A;Reference number: I47140; MUID:95080454; PMID:7988744
A;Accession: I47140
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-595 <BOK>
A;Cross-references: EMBL:Z37167; NID:g587554; PIDN:CAA85524.1; PID:g587555
A;Experimental source: uterus
R;Thole, H.H.; Maschler, I.; Junqblut, P.W.
Eur. J. Biochem. 231, 510-516, 1995
A;Title: Surface mapping of the ligand filled C-terminal half of the porcine estradio
A;Reference number: S66250; MUID:95361877; PMID:7645164
A;Accession: S66250
```

```
A;Molecule type: protein
A;Residues: 297-307;310-313;320-323;329-332;337-340;417-420;466-473 <THO>
A;Experimental source: uterus
R;Thole, H.H.
FEBS Lett. 320, 92-96, 1993
A;Title: Assignment of the ligand binding site of the porcine estradiol receptor to the
A;Reference number: S32402; MUID:93269381; PMID:8458437
A;Accession: S32402
A;Molecule type: protein
A;Residues: 303-423 <THW>
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; phosphoprotein; steroid hormone receptor; transcriptio
F;1-120/Domain: amino terminal <NH2>
F;121-299/Domain: DNA binding #status predicted <DNA>
F;183-456/Domain: erbA transforming protein homology <ERBA>
F;185-205/Region: zinc finger CCCC motif
F;221-245/Region: zinc finger CCCC motif
F;256-271/Region: nuclear location signal
F;400-595/Domain: steroid binding #status predicted <STR>
F;185,188,202,205/Binding site: zinc (Cys) #status predicted
F;221,227,247,240/Binding site: zinc (Cys) #status predicted
F;236,295/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             43.5%;  Score 1219.5;  DB 2;  Length 595;
  Best Local Similarity   46.0%;  Pred. No. 2.6e-87;
  Matches  268;  Conservative   89;  Mismatches  126;  Indels   99;  Gaps   19;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSSYVDSHHEYPAMTFYSPAVMNY------ 56
              |    || |      |  :|||   | :|: ||                 ||| ||
Db         21 NELEPLNRP----  QLKIPLERFLGEVYVDSS ------         KEAVENYTEGAAT 69

Qy         57       --SIP SNVTNLEGGPGRQ          -----      TTSPNVLWPTPGHLSP- 88
                      | |    : |  |||  :                    : || ||   |  ||| 
Db         61 DFNAAAAASAPVYGQSGLAYGPGSEAAAFGANGLGGFQPLNSVSPSPLVLIHPPPQLSPF 120

Qy         89 LVVHRQLSHLYAFPQKSPWC--FARSLEHTLPVNRETLKRKVSGNRCASPVTGPG----- 141
              | | |    | : | :  ||       |    :  |: |    :  :  |
Db        121 LHPHGQQVPYYLENEPSGYAVREFAGPPAFYRP---NSDNRPQGGRERLASTSDKGSMAME 177

Qy        142 SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSC 201
              | ::  :|||:|||||||||||||||||||||||||||||:|||||||||||||||||
Db        178 SAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSC 237

Qy        202 QACRLRKCYEVGMVKGGSRKERGGYRLVRKQKSADEQLHCACKAK---------KSQG-- 250
              |||||||||||||:| | |::| | |::: :|  |:     |: :          ||
Db        238 QACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRNEAVPPCGMRSANIEW 292

Qy        251 -------HAPRVRELLLDALSPEQLVLTLLEAEPPHVLISR  PSAPFTEASMMMSLTKL 301
                    |  :   :|  :|: :|::  ||||||| :: |   |: | :|||||  || |
Db        293 PSPLLIKHTKKNSPVL  SLTADQMISALLEAEPPP-IIYSEYDPTRPFSEASMMGLLTNL 349

Qy        302 ADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLV 361
              |:|||||||:|||::|||::||| ||| |||  |:|:||||||||||||||||||||||
Db        350 ADRELVHMINWAKRVPGFIDLSLHDQVHLLECAWLEILMIGLVWRSMEHPGKLIFAPNLL 409

Qy        362 LDRDMGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 421
              |||::|||||||::|||||||||||:|||| : || :|:|:|:|:|||| :|  :::| :
Db        410 LDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSL 469

Qy        422 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 481
              :    :  :|| |: ::|||:|: |||  ||| ||:||| || |||||||| |||
Db        470 EEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMK 529

Qy        482 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSEGSPALDSKSKESQNPQS 529
              ||||||:|||||||||:|| |    ::: |    | ||    :
Db        530 CKNVVPLYDLLLEMLDAHRLHAPTSRGG  EPPEDMSQSQ 567


RESULT 7
QRMSE
estrogen receptor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Mar-1992 #sequence_revision 14-Jul-1994 #text_change 22-Jun-1999
C;Accession: A40061
R;White, R.; Lees, J.A.; Needham, M.; Ham, J.; Parker, M.
Mol. Endocrinol. 1, 735-744, 1987
A;Title: Structural organization and expression of the mouse estrogen receptor.
A;Reference number: A40061; MUID:91042558; PMID:2484714
A;Accession: A40061
A;Molecule type: mRNA
A;Residues: 1-599 <WHI>
A;Cross-references: GB:M38651; NID:g193179; PIDN:AAA37580.1; PID:g193180
C;Comment: The steroid hormones and their receptors are involved in the regulation of
C;Comment: In the absence of ligand, steroid hormone receptors are thought to be weak
complex appears to recognize discrete DNA sequences upstream of transcriptional start
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; phosphoprotein; steroid binding; steroid hormone re
F;1-183/Domain: amino terminal <NH2>
F;184-275/Domain: DNA binding #status predicted <DNA>
F;187-460/Domain: erbA transforming protein homology <ERBA>
F;187-210/Region: zinc finger CCCC motif
F;223-245/Region: zinc finger CCCC motif
F;260-275/Region: nuclear location signal
F;304-556/Domain: steroid binding #status predicted <STR>
F;189,192,206,209/Binding site: zinc (Cys) #status predicted
F;225,231,241,244/Binding site: zinc (Cys) #status predicted
F;240,300/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             42.4%;  Score 1218.5;  DB 1;  Length 599;
  Best Local Similarity   45.8%;  Pred. No. 3.2e-87;
  Matches  264;  Conservative   98;  Mismatches  152;  Indels   61;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|:|   | :|: :|   | :: |  |  | : |    : :
Db         21 NELEPLNRP------QLKMPMERALGEVYVDNSKPTVFNYPEGAAYEFNAAAAAAAASA 74

Qy         61 NVTNLEG---GPGRQ-----------------TISPNVLWPTPGHLSP-LVVHRQLSHL 98
              |    |   |||  :                 : || :|   |  ||| |  | |
Db         75 PVYGQSGIAYGPGSEAAAFSANSLGAFPQLNSVSPSPLMLIHPPPQLSPFLHPHGQQVPY 134

Qy         99 YAEPQKSPWCEARSLEHTLP---VNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCA 150
              | | : |    | ::  | |       :  |: :|    |     |    | ::  :||
Db        135 YLENEPS----AYAVREFAGPPAFYRSNSDNRPQNGRERLSSSNEKGNMIMESAKETRYCA 190

Qy        151 VCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCY 210
              |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        191 VCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCY 250

Qy        211 EVGMVKGGSRKERGGYRLVRKQKSADE-----QLHCACKAKRSQ--------GHAPPVRE 257
              |||||| | |::| | |::: :|  |:      ::  :|  : :           | :
Db        251 EVGMMKGGIRKDRRGGRMLKHKRQRDDLEGRNEMGASGDMRAANLWPSPLVIKHTKKNSP 310

Qy        258 LLLDALSPEQLVLTLLEAEPPHVLISR-PDSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :      || ||:|||||  || |||:|||||::|||::
Db        311 AL--SLTADQMVSALLDAEPPMIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRV 368

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDMGKCVEGILEI 376
              ||| :|:| ||| |||  |:|:||||:|||||:|||||||||||||:|||::|||||:||
Db        369 PGFIDLSLHDQVHLLECAWLEILMIGLVWRSMEHPGKLIFAPNLLLDRNQGKCVEGMVEI 428

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||||| : || :|:|:|:|||| :|  :::| :   :    : :|: :||
Db        429 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 488

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::|||:|:: |||  ||| ||::|||:|| |||||||| |||||||||:|||||||
Db        489 TLIHLMAKAGLTLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEML 548

Qy        496 NAHVLRGCKSSITGSEGSPALDSKSKESQNPQS 529
              :|| |    | :      |::   :   | :  |
Db        549 DAHRLHAPASRMGVPPEEPSQTQLATTSSISAHS 582


RESULT 8
QRRTE
estrogen receptor - rat
```

```
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1991 #sequence_revision 14-Jul-1994 #text_change 22-Jun-1999
C;Accession: S07379; S16731
R;Koike, S.; Sakai, M.; Muramatsu, M.
Nucleic Acids Res. 15, 2499-2513, 1987
A;Title: Molecular cloning and characterization of rat estrogen receptor cDNA.
A;Reference number: S07379; MUID:87174780; PMID:3031601
A;Accession: S07379
A;Molecule type: mRNA
A;Residues: 1-600 <KOI>
A;Cross-references: EMBL:Y00102; NID:g56110; PIDN:CAA68287.1; PID:g56111
R;Maggi, A.M.A.
submitted to the EMBL Data Library, June 1991
A;Reference number: S16731
A;Accession: S16731
A;Molecule type: mRNA
A;Residues: 1-487,'T',489-600 <MAG>
A;Cross-references: EMBL:X61098; NID:g56120; PIDN:CAA43411.1; PID:g56121
C;Comment: The steroid hormones and their receptors are involved in the regulation of eu
C;Comment: In the absence of ligand, steroid hormone receptors are thought to be weakly
omplex appears to recognize discrete DNA sequences upstream of transcriptional start sit
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; phosphoprotein; steroid binding; steroid hormone recep
F;1-184/Domain: amino terminal <NB2>
F;185-276/Domain: DNA binding #status predicted <DNA>
F;188-461/Domain: erbA transforming protein homology <ERBA>
F;188-211/Region: zinc finger CCCC motif
F;224-246/Region: zinc finger CCCC motif
F;261-276/Region: nuclear location signal
F;305-557/Domain: steroid binding #status predicted <STB>
F;190,193,207,210/Binding site: zinc (Cys) #status predicted
F;226,232,242,245/Binding site: zinc (Cys) #status predicted
F;241,310/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             43.4%;  Score 1216.5;  DB 1;  Length 600;
  Best Local Similarity   45.7%;  Pred. No. 4.5e-87;
  Matches  262;  Conservative   99;  Mismatches  145;  Indels   77;  Gaps   17;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSSYVDSHHEYPAMTPY-    - SPAVMNY 56
              |   || |      |  :||:|   | :|: :|   :   ||    |       : |
Db         21 NELEPLNRP- ---- QIKMPMPRALGFVYVDNS-KPAVFNYPFGAAYFFNAAAAAAAAGA 78

Qy         57 STP----SNVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSPLV---V 91
              | |    |:::|    |||  :                 : || :|   | |:|| :
Db         74 SAPVYGQSSITY---GPGSEAAAFGANSLGAFPQLNSVSPSPLMLLHPPPHVSPFLHPHG 130

Qy         92 HRQLSHLYAEPQKSPWCEARSLEHTLP---VNRETLKRKVSGNRCASPVTGPG-----SK 143
              |:  :|  ||       | ::  | |       :  |: :|   |  :  |       |
Db        131 HQVPYYYLENEP------SAYAVRDTGPPAFYRSNSDNRRQNGRERLSSSSEKGNMIMESA 184

Qy        144 RDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQA 203
              ::  :|||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db        185 KETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQA 244

Qy        204 CRLRKCYEVGMVKGGIRKDRRGGRMLKHKRQRDDG-PGRNEMGESLDMRAANLWPSPLVIK 262
              |||||||||||:| | |::| | ):::  :|  |:      ::  :|  : :
Db        245 CRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEPGRNEMGESLDMRAANLWPSPLVIK 304

Qy        263 HAIEKVRELLDALSPEQIVLILLEAFEPHVLIS-RPSAPFTEASMMMSLTKLADKELVHM 321
              |  :    |  :|: :|:|  ||:|||| :     || ||:|||||  || |||:|||||
Db        305 HTKKNSPALT--SLTADQMVSALLDAEPPLIYSEYDPSRPFSEASMMGLLTNLADRELVHM 362

Qy        322 ISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKC 381
              |:|||::||  :|:|  ||| |||   |:|:||:|||:::|||||:|||::|||:|||||
Db        363 INWAKRVPGFGDLNLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKC 422

Qy        382 VEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAH 440
              |||::|||||||||:|||| : || :|:::|:|::|||||  :|   :::|  :   :
Db        423 VEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHR 482

Qy        441 LLNAVTDALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVY 500
              :|: : | |: ::||:|:: |||  ||| ||::|||:|| ||||||| ||||||||||:|
Db        483 VLDKINDTLIHLMAKAGLTLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLY 542

Qy        501 DLLLEMLNAHVLRGCKSSITGSECSPAEDSKSK 533
              |||||||:|| |    | :      | | |:|:
Db        543 DLLLEMLDAHRLHAPASRM---GVPPEEPSQSQ 572


RESULT 9
ORXLE
estrogen receptor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 28-Feb-1992 #sequence_revision 14-Jul-1994 #text_change 15-Jun-1996
C;Accession: A40907
R;Weiler, I.J.; Lew, D.; Shapiro, D.J.
Mol. Endocrinol. 1, 355-362, 1987
A;Title: The Xenopus laevis estrogen receptor: sequence homology with human and avian
A;Reference number: A40907; MUID:90331927; PMID:3274894
A;Accession: A40907
A;Molecule type: mRNA
A;Residues: 1-586 <WEI>
A;Cross-references: GB:L20735
C;Comment: The steroid hormones and their receptors are involved in the regulation of
C;Comment: In the absence of ligand, steroid hormone receptors are thought to be weak
omplex appears to recognize discrete DNA sequences upstream of transcriptional start
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; phosphoprotein; steroid binding; steroid hormone re
F;1-174/Domain: amino terminal <NB2>
F;175-266/Domain: DNA binding #status predicted <DNA>
F;178-448/Domain: erbA transforming protein homology <ERBA>
F;178-201/Region: zinc finger CCCC motif
F;214-237/Region: zinc finger CCCC motif
F;251-266/Region: nuclear location signal
F;292-544/Domain: steroid binding #status predicted <STB>
F;180,184,197,200/Binding site: zinc (Cys) #status predicted
F;216,222,232,235/Binding site: zinc (Cys) #status predicted
F;231/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             43.3%;  Score 1214;  DB 1;  Length 586;
  Best Local Similarity   48.6%;  Pred. No. 6.9e-87;
  Matches  255;  Conservative   82;  Mismatches  138;  Indels   50;  Gaps   12;

Qy         35 YVDSHH----EYP----------AMTFYSPAVMNYSTPSNVTNLEGGPGPQF-----TSP 76
              ||:::     ||          |  || | ::|:  |         |  |      ||
Db         43 YVENNRTCTFNYPECTTYDFAAAAAPVYSSASLSYAASSFTPGSSSLTGLHTLNNVPPSP 102

Qy         77 NVLWPTPGHLSPLVVHR-QLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCAS 135
              |       ||| : | |   | | ::  :    :  |      :  |: ||    |
Db        103 VVFLAKLPQLSPFIHHHGQQVPYYLESEQGTFAVPEAAPPTF-YPSSSDNRPQSGRERMS 161

Qy        136 PVTGPG-----SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
              |     | ::  :|||||||||||||||||||||||||||||||||||||:||||||
Db        162 SANDKGPPSMESTKETRYCAVCSDYASGYHYGVWSCPGCKAFFKRSIQGHNDYMCPATNQ 221

Qy        191 CTIDKNRRKSCQACRLRKCYEVGMVKGCSRKERGGYRTVPRQPSADPQLHCAGKAKPSGG 250
              |||||||||||||||||||||||||:| | |::| | ||:: :|  :||     : |
Db        222 CTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRLLKHKRQRDEQ     LQKNDVP 276

Qy        251 HA----------PPVRELLED---ALSPEQLVLTLLEAEPPHVLISRPSA-PFTEASMMM 296
              :           | |: : |    :|: |||:  |:||| | |     |  | :|||||
Db        277 PSEIRTASIWVNPSVKSMKLSPVLSLIAEQLISALMEAEAPIVYSEHDSTKPLSEASMMT 336

Qy        297 SLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIF 356
              || |||:||||||:|||::||||:|:| ||| |||  |:|:||:|||:::|||||:|||| |
Db        337 LLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMVGLIWRSVEHPGKLSF 396

Qy        357 APDLVLDRDEGKCVEGILEIFDMLLATTSREFELKLQHKEYLCVKAMILLNSSMYPLVTA 416
              ||:|:|||::|:||||:|||||: | |||| ::|: :|:|:|::|||||| :|  :::
Db        397 APNLLLDRNQGRCVEGLVEIFDMLVTTATRFRMMRLRGEEFICLKSIILLNSGVYTFLSS 456

Qy        417 TQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEH 475
              | ::     | | :|: : | ||  :||||:| |||  ||| ||::|||:|| ||||||
Db        457 TLESLEDHDLIHILDKIIDTLVHFMAKSGLSLQQQQRRLAQLLLILSHIRHMSNKGMEH 516

Qy        476 LLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKS 520
              |||||||||||||||||||||||||||||||||||||||||||||
Db        517 LLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKS 561
```

```
                  | :|||||||||:|||||||||:|| :   |   |  |    |||:|
Db        517 LYSMKCKNVVPLYDLLLEMLDAHRIHTPKDKTTIQE----EDSRS 557


RESULT 10
T10423
estrogen receptor - Oryzias sp. (strain d-rR)
C;Species: Oryzias sp
A;Variety: strain d-rR
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Jun-2000
C;Accession: T10423
R;Okada, H.; Kawahara, I.; Yamashita, I.
submitted to the EMBL Data Library, March 1994
A;Description: Cloning of medaka estrogen receptor cDNA.
A;Reference number: Z17013
A;Accession: T10423
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-620 <OKA>
A;Cross-references: EMBL:D28954
A;Experimental source: strain d-rR, liver
C;Genetics:
A;Gene: MER
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation
F;184-460/Domain: erbA transforming protein homology <ERB>

  Query Match             42.7%;  Score 1199;  DB 2;  Length 620;
  Best Local Similarity   48.1%;  Pred. No. 1.1e-85;
  Matches  265;  Conservative   81;  Mismatches  129;  Indels   76;  Gaps   16;

Qy          36 YVDSHHEY----PAMT---------FYSPAVMNYSIPS--NVTNLEGGPGRQTTSPNVLW 80
               :::  ::|    || |          :|| :     ||  :: :|  ||    ||| |
Db          62 FLEGTYDYAAPNPATTPLYSQSSTGYYSAPLETNGPPSEGSLQSLGSGP----TSPLVEV 117

Qy          81 PTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNPETLKRKVSGNR---CASPV 137
               |:   ||| :     || | |   :|   :          ::    :    |:|   |:
Db         118 FSSPRLSPFM--HPPSHHYLETTSTPVYKSS--------HQGASREDQCGSREDTCSLGE 167

Qy         138 TGPGS-------KRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
               | |:              :|  |||||||||||||||||||||||||||||||||:||||||
Db         168 LGAGAGAGGFEMAKDTRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQ 227

Qy         191 CTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR--------------SAD 236
               ||||:|||||||||||||||||:| | |::|   |::|| :
Db         228 CTIDRNRRKSCQACRLRKCYEVGMMKGGVRKDR--IRILRRDKRPTGVGDGDEVVKGQPH 285

Qy         237 EQLHCAGKAKRS--GGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP--SAPFTEA 292
               : :|  |: : |  ||        | : :: |||::| |  |||| :| ||   | |:||
Db         286 KTVHYDGRKRSSTGGGGGGGGGRLSVTSIPPEQVLLLLQGAEPP-ILCSPQKLSKPYTEV 344

Qy         293 SMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG 352
               :||  || :|||||||||:||||:|||:||| ||| |||| |:||||:|:|||:|||  ||
Db         345 IMMTLLTSMADKELVHMIAWAKKLPGFIQLSLHDQVLLLESSWLEVLMIGLIWRSIHCPG 404

Qy         353 KLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYP 412
               ||||| ||:|||:|| |||| |||||||||| |||| ||| :|:|:|:|:|||||  :
Db         405 KLIFAQDLILDRNEGDCVEGMTEIFDMLLATASRFRVLKLKPEEFVCLKAIILLNSGAFS 464

Qy         413 LVTATQD-ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNK 471
               | | :   :|  :  :|: :||||:  |::||  :|:|: | | ||:|||:|| |||
Db         465 FCTGTMEPLHNSAAVQSMLDTITDALIHYISQSGYLAQEQARRQAQLLLLLSHIRHMSNK 524

Qy         472 GMEHLLNMKCKNVVPVYDLLLEMLNAHVL----RGCKS----------SITGSEGSPAE 516
               ||||| :||||| ||:|||||||||:|| |   |  :|       | |   :|
Db         525 GMEHLYSMKCKNKVPLYDLLLEMLDAHRLHHPVRAPQSLSQVDRDPPSTSSGGGGIAPGS 584

Qy         517 DSKSKEGSQNP 527
                 | |:   ::|
Db         585 ISASRGRIESP 595


RESULT 11
```

```
S64737
80K estrogen receptor - human
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1996 #sequence_revision 13 Mar 1997 #text_change 20-Aug-1999
C;Accession: S64737
R;Pink, J.J.; Wu, S.Q.; Wolf, D.M.; Bilimoria, M.M.; Jordan, V.C.
Nucleic Acids Res. 24, 962-969, 1996
A;Title: A novel 80 kDa human estrogen receptor containing a duplication of exons 6 a
A;Reference number: S64737; MUID:96174665; PMID:8600466
A;Accession: S64737
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-701 <PIN>
A;Cross-references: EMBL:U47678; NID:g1197854; PIDN:AAB00115.1; PID:g1197855
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: steroid hormone receptor; zinc finger
F;182-456/Domain: erbA transforming protein homology <ERBA
F;518-562/Domain: erbA transforming protein homology #status atypical <ERB2

  Query Match             41.9%;  Score 1174.5;  DB 2;  Length 701;
  Best Local Similarity   41.2%;  Pred. No. 1.1e-83;
  Matches  268;  Conservative   87;  Mismatches  125;  Indels  171;  Gaps   16;

Qy           5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMIFYSPAVMNYSIPS 60
               |   || |      |  :|||   | :|: ||   | :: |  |  | : |  |  :
Db          21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy          61 NVINLEGGPGRQ-----------------ITSPNVLWPIPGHLSP-LVVHRQLSHLYAE 101
                 | |  ||| :                 : || :|   | ||| |  | |    | |
Db          75 Q-TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy         102 PQKSPWC--EAPSLEHTLPVNPETLKRFVSGNRCASPVTGPG-----SKPDAHFCAVCSD 154
                 : |  :   ||        |    :  |:  |    :       |     | ::  :|||:|
Db         134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy         155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
               ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db         191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy         215 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVR----------------EL 258
               :| | |::| | |::: :|  |:     |: :   | |  :|
Db         251 MKGGIRKDRRGGRMLKHKRQRDD-----GEGPGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy         259 LLDALSPEQLVLTLLEAEPPHVLISR--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
               |  :|: :|:|  |||:||| :| |   |: |||:|||  || |||:|||||:|:::
Db         306 LALSLTADQMVSALLDAEPP-ILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy         317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
               ||||:|:| ||| |||  |:|:|||:|:|||:|| ||:|||||:|||:|||||||||:||
Db         365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy         377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT------QD---------- 419
               |||||||:|||| : || :|:|:|:|:|||| :|  :::|      :|
Db         425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy         420 ----------------------------------------------------ADSSR---- 424
                                                                   | |||
Db         485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNQGMEHLYSMKCKNVVPLYDLLLEMLN 544

Qy         425 ----------------------------------KLAHLLNAVTDALVWVIAKSGISSQ 449
                                                 :  :|: :|| |: :|||:|: |
Db         545 NLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQ 604

Qy         450 QQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
               ||  ||| ||:|||:|| |||||||| :|||||||||:|||||||:|| |
Db         605 QQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRL 655


RESULT 12
S58224
oestrogen receptor - Atlantic salmon (fragment)
C;Species: Salmo salar (Atlantic salmon)
```

```
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
C;Accession: S58224
R;Rogers, S.A.; Llewellyn, L.; Ramsurn, V.P.; Sweeney, G.E.; Wigham, T.
submitted to the EMBL Data Library, July 1995
A;Reference number: S58224
A;Accession: S58224
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-545 <ROG>
A;Cross-references: EMBL:X89959; NID:g929901; PIDN:CAA61999.1; PID:g929902
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: steroid hormone receptor; zinc finger
F;103-382/Domain: erbA transforming protein homology <ERBA>

  Query Match             40.4%;  Score 1134.5;  DB 2;  Length 545;
  Best Local Similarity   48.8%;  Pred. No. 1e-80;
  Matches  246;  Conservative   73;  Mismatches  114;  Indels   71;  Gaps   15;

Qy          70 GRQTTSPNVLWPTPGHLSPLVVHRQLSH----  LYAEPQKSPWCEARSLEHTLPVNREI 124
               |   | | |    :   |||  :|    |       | |   :|    :  : : |  : |
Db          17 GSSPTGPLVFVSSSPQLSPF-LHPPGHHGLPSQSYYLETSSTPLYKSSVVTNQLSASFF- 74

Qy         124 LKRKVSGNRCAS-PVTGPGSK - --RDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSI 177
               |  :: ||  |    | | :       :  :|||||:|||||||  ||||||||||||||
Db          75  KLCITSNRQQSYAAACSGVPVFFMANFTRYCAVCSDFASGYHYGEWSCEGCKAFFKRSI 133

Qy         178 QGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERCGYRLVRRQKSADE 237
               |||||:|||||||:|:||||||||||||||||||| | |::| | |::|:    :
Db         134 QGHNDYMCPATNQCTMDRNRRKSCQACRLRKCYEVGMVKGGLRKDR-GGRVLRK- -- D 187

Qy         238 QLHC--AGKAKR - -  -  ---- -- --  --SGGH---APRVRELLLDALSPEQL 288
               : :|  ||  ::                    |||     ||:       : |||:
Db         188 KRYCGPAGDREKPYGDLEHRTAPPQDGGRNSSSSSLSCGGWCGPP!--- --TMPPPQV 241

Qy         269 VLTLLEAEPPHVLISRP- SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFD 326
               :  |  ||||  | ||   : |:|| :||  || :|||||||||:||||:||| |||| |
Db         242 LFLLQGAEPP ALCSRQKVARPYTEVTMMTLLTSMADKELVHMIAWAKKVPGFQELSLHD 300

Qy         327 QVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSR 386
               ||:|||| |:||||:||:|||||  |||||| ||:||| || |||| ||||||||| ||
Db         301 QVQLLESSWLEVLMIGLIWRSIHCPGKLIFAQDLILDRSEGDCVEGMAEIFDMLLATVSR 360

Qy         387 FRELKLQHKEYLCVKAMILLNSSMYPLVT-ATQDADSSRKLAHLLNAVTDALVWVIAKSG 445
               || |||: :|:::|:||:|||||  :   : : :   :|  :  :|: :||||:  |: ||
Db         361 FRMLKLKPEEFVCLKAIILLNSGAFSFCSNSVESLHNSSAVFSMLDNITDALIHHISHSG 420

Qy         446 ISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKS 505
               | |||  |   ||:|||:|| |||||||| ::|||| ||:|||||||: | |:
Db         421 ASVQQQPRRQVQLLLLLSHIRHMSNKGMEHLYSIKCKNKVPLYDLLLEMLDGHRLQ -   476

Qy         506 SITGSEGSPAEDSKSKEGSQNPQS 529
                  || : ::: | :: | :
Db         477    -----SPGKVAQAGEQTEGPST 498


RESULT 13
A37197
estrogen receptor - rainbow trout
C;Species: Oncorhynchus mykiss (rainbow trout)
C;Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 20-Aug-1999
C;Accession: A37197; A40070
R;Pakdel, F.; Le Gac, F.; Le Goff, P.; Valotaire, Y.
Mol. Cell. Endocrinol. 71, 195-204, 1990
A;Title: Full length sequence and in vitro expression of rainbow trout estrogen receptor
A;Reference number: A37197; MUID:91006824; PMID:2210841
A;Accession: A37197
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-574 <PAK>
A;Cross-references: GB:M31559
R;Pakdel, F.; Le Guellec, C.; Vaillant, C.; Le Roux, M.G.; Valotaire, Y.
Mol. Endocrinol. 3, 44-51, 1989
A;Title: Identification and estrogen induction of two estrogen receptors (ER) messenger
A;Reference number: A40070; MUID:89127284; PMID:2915648
A;Accession: A40070
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 150-574 <PA2>
A;Cross-references: GB:M31559; NID:g213783; PIDN:AAA49552.1; PID:g213784
C;Superfamily: estrogen receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation;
F;145-421/Domain: erbA transforming protein homology <ERBA>
F;147-167/Region: zinc finger
F;183-207/Region: zinc finger

  Query Match             39.9%;  Score 1120;  DB 2;  Length 574;
  Best Local Similarity   48.8%;  Pred. No. 1.5e-79;
  Matches  243;  Conservative   70;  Mismatches  127;  Indels   58;  Gaps   12;

Qy          70 GRQTTSPNVLWPTPGHLSPLV---VHRQLSH------LYAEPQKSPWCFARSLEHTLPVN 120
               |   | | |    :   ||| :   :|    |       | |   :|    :  : : |
Db          55 GSSPTGPLVFVSSSPQLSPQLSPFLHPPSHHGLPSQSYYLETSSTPLYRSSVVTNQLSAA 114

Qy         121 REI-----ERKVSGNRCASPVIGPGSKEDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRS 176
               | |     |: | :   | |       :  :|||||:|||||||||||||||||||||
Db         115 EEKLCIASDRQQSYSAAGSGVRVFEMANELRYCAVCSDFASGYHYGVWSCEGCKAFFKRS 174

Qy         177 IQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERCGYRLVRR---- 231
               |||||||:|||||||:|:|||||||||||||||||| | |::| | |::|:
Db         175 IQGHNDYMCPATNQCTMDRNRRKSCQACRLRKCYEVGMVKGGLRKDR-GGRVLRKDKRYC 233

Qy         232 ----  -          --QRSADEQLHCAGKAKRSGG-HAPRVRELLLDALSPEQLVLTLLE 274
                               || |   : :      ||   ||:       : ||| || ||:
Db         234 GPAGDREKPIVTWSTGQRPQDGGRNSSSSLNGGGGWPGPRI- ---TMPPEQ VLFLLQ 286

Qy         275 AEPPHVLISRP- SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLE 332
               : |  | ||   : |:|| :||  || :|||||||||:||||:||| |||| |||:|||
Db         287 GQIP-ALCSRQKVARPYTEVTMMTLLTSMADKELVHMIAWAKKVPGFQELSLHDQVQLLE 345

Qy         333 SCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL 392
               | |:|||:||:|||  |||||| ||:||| || |||: |||||||| |||  |||
Db         346 SSWLEVLMIGLIWRSIHCPGKLIFAQDLILDRSEGDCVEGMAEIFDMLLATVSRFGMLKL 405

Qy         393 QHKEYLCVKAMILLNSSMYPLVT-ATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQ 451
               : :|::|:||:|||  :   : : :   :|  :  :|: :||||:  |: || | |||
Db         406 KPEEEFVCLKAILLLNPGAFSFCSNSVESLHNSSAVESMLDNITDALIHHISHSGASVQQQ 465

Qy         452 SMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSE 511
               | | ||:|||:|| |||||||| ::|||| ||:|||||||: | |:
Db         466 PRRQAQLLLLLSHIRHMSNKGMEHLYSIKCKNKVPLYDLLLEMLDCHRLQ- -- - --  515

Qy         512 GSPAEDSKSKEGSQNPQS 529
               :| : ::: | :: | :
Db         516 -APGKVAQAGEQTEGPST 532


RESULT 14
B29345
steroid hormone receptor ERR2 precursor - human
C;Species: Homo sapiens (man)
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Sep-1999
C;Accession: B29345
R;Giguere, V.; Yang, N.; Segui, P.; Evans, R.M.
Nature 331, 91-94, 1988
A;Title: Identification of a new class of steroid hormone receptors.
A;Reference number: A29345; MUID:88122546; PMID:3267207
A;Accession: B29345
A;Molecule type: mRNA
A;Residues: 1-433 <GIG>
A;Cross-references: GB:X51417; NID:g36610; PIDN:CAA35779.1; PID:g36611; GB:Y00290
C;Superfamily: unassigned erbA related proteins; erbA transforming protein homology
C;Keywords: DNA binding; steroid hormone receptor; transcription regulation; zinc fin
F;101-352/Domain: erbA transforming protein homology <ERBA>
F;103-123/Region: zinc finger
F;139-163/Region: zinc finger
```

```
  Query Match             22.5%;  Score 631.5;  DB 2;  Length 433;
  Best Local Similarity   39.4%;  Pred. No. 1.5e-41;
  Matches  142;  Conservative   69;  Mismatches  110;  Indels   40;  Gaps    9;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | || | ||||||||| ||| |||||||:||: :| ||||||| | | |||||||||  |
Db        103 CLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPAINECEITKRRRKSCQACRFMK 162

Qy        209 CYEVGMVKGGSR--RERGGYRLVRRQRSADFQLHCAGKAKPSGGHAPPVRELLLDALSPE 266
              | :|||:| | |  | | | :  :|:  ::   :                  | |    |
Db        163 CLKVGMLKEGVRLDRVRGGRQKYKRRPIDSPNSPY------------     LPTQTSPPA 205

Qy        267 QLVLT     LLEAEPPHVLISRP-SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
              :  ||     || |||  :    |   |  :   : :|  |||:||| :|||| ||||
Db        206 KRPITKIVSYIIVAEPDKLYAMPPDDVPEGDIKALTTLCDLADRELVFLISWAKHIPGFS 265

Qy        321 ELSLEDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
              |:| ||: |||| |||:|::|:::||: :  || :| | ::| :  : | |::|::  :
Db        266 NLTLGDQMSLLQSAWMEILILGIVYRSLPYDDKLAYAEDYIMDEEHSRLV-GLLDLYPAI 324

Qy        381 LATTSRFRELKLQHKEYLCVKAMILLNS-SMYPLVTATQDADSSRKLAHLLN-AVTDALV 438
              |    |:::||:: :||: :||: | || |||      :: :: :|| ||: |: | :
Db        325 LQLVPRYKKLKVEKEEFVMLKALALANSDSMY-----IENLEAVQKLQDLLHEALQDYEL 379

Qy        439 WVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
                        ::  |   || : |  :|  : | ::|  ::| :  || : | |||  |
Db        380 -------SQRHEEPRRAGKLLLTLPLLRQTAAKAVQHFYSVKLQGKVPMHKLFLEMLEAK 432

Qy        499 V 499
              |
Db        433 V 433


RESULT 15
S58087
estrogen receptor related protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Sep-1999
C;Accession: S58087
R;Pettersson, K ; Svensson, K ; Mattsson, R ; Carlsson, B ; Ohlsson, P.; Berkenstam, A.
submitted to the EMBL Data Library, July 1995
A;Description: Expression of a novel member of estrogen response element-binding nuclear
A;Reference number: S58087
A;Accession: S58087
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-433 <PET>
A;Cross-references: EMBL:X89594; NID:g914855; PIDN:CAA61755.1; PID:g914856
C;Superfamily: unassigned erbA-related proteins; erbA transforming protein homology
C;Keywords: steroid hormone receptor; zinc finger
F;101-352/Domain: erbA transforming protein homology <ERBA>

  Query Match             22.1%;  Score 619.5;  DB 2;  Length 433;
  Best Local Similarity   38.7%;  Pred. No. 3.7e-40;
  Matches  139;  Conservative   76;  Mismatches  108;  Indels   36;  Gaps   10;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | || | ||||||||| ||| |||||||:||: :| ||||||| | | |||||||||  |
Db        103 CLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYNCPAINECEITKRRRKSCQACRFMK 162

Qy        209 CYEVGMVKGGSR--RERGGYRLVRRQRSADFQ-  QLHCAGKAKPSGGHAPPVRELLLDA 262
              | :|||:| | |  | | | :  :|:  ::     |  :|  |||:       : :  :
Db        163 CLKVGMLKEGVRLDRVRGGRQKYKRRPIDSPNSPYLNLPISPPAKKP -----LIKIVGNL 216

Qy        263 LSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVEL 322
              | ||      | | |||:         |  :    : :| :|||:||| :|:|||| |||| '
Db        217 LGVEQ---DKLYAMPPN------DIPEGDIKALTTLCELADRELVFLINWAKHIPGFPSL 267

Qy        323 SLEDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA 382
              :| ||: ||:| |||:|::|:::||: :  || :| | ::| :  : | |::|::  :|
Db        268 TLGDQMSLLQSAWMEILILGIVYRSLPYDDKLAYAEDYIMDEEHSRLV-GLLDLYPAILQ 326

Qy        383 TISRFRELKLQHKEYLCVKAMILLNS-SMYPLVTATQDADSSRKLAHLLN-AVTDALVWV 440
                                                                  ||
Db        327 LVRKYKKLKVEKEEFMILKALALANSDSMY-----IENLEAVQKLQDLLHCALQDYEL-- 379

Qy        441 IAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHV 499
                        ::  |   || : |  :|  : | ::|  ::| :  || : | |||  | |
Db        380 -----SQRHEEPRRAGKLLWTLPLLRQTAAKAVQHFYSVKLQGKVPMHKLFLEMLEAKV 433


RESULT 16
A29345
steroid hormone receptor ERR1 precursor - human
N;Alternate names: estrogen-related receptor
C;Species: Homo sapiens (man)
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Sep-1999
C;Accession: A29345; A49074
R;Giguere, V.; Yang, N.; Segui, P.; Evans, R.M.
Nature 331, 91-94, 1988
A;Title: Identification of a new class of steroid hormone receptors.
A;Reference number: A29345; MUID:88122546; PMID:3267207
A;Accession: A29345
A;Molecule type: mRNA
A;Residues: 1-521 <GIG>
A;Cross-references: EMBL:X51416; NID:g36608; PIDN:CAA35779.1; PID:g36609; EMBL:Y08290
R;Wiley, S.R.; Kraus, R.J.; Zuo, F.; Murray, E.E.; Loritz, K.; Mertz, J.E.
Genes Dev. 7, 2206-2219, 1993
A;Title: SV40 early-to-late switch involves titration of cellular transcriptional rep
A;Reference number: A49074; MUID:94040741; PMID:8224847
A;Accession: A49074
A;Status: preliminary
A;Molecule type: protein
A;Residues: 166-169,'X',171-173 <WIL>
C;Superfamily: unassigned erbA related proteins; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation;
F;174-434/Domain: erbA transforming protein homology <ERBA>
F;176-196/Region: zinc finger
F;212-236/Region: zinc finger

  Query Match             21.3%;  Score 597.5;  DB 2;  Length 521;
  Best Local Similarity   33.8%;  Pred. No. 8.8e-39;
  Matches  185;  Conservative   68;  Mismatches  211;  Indels   83;  Gaps   19;

Qy          1 MDIKNSPSSLN--------SPSSYNC----------SQSILPLE---HGSIYIPSSYVD 38
              |  |:|| ||:        | |  |           |:   | |    |   |    :
Db          5 MSSKDSPGSLDGRAWEDAQKPQSAWCGGRKIPVYATSSRRAPPSEGIERGGAARPE---E 61

Qy         39 SHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTP---GHLSPLVVHRQL 95
              :  | |      :|  : :|          ||      | | | |      :|  ||  :
Db         62 AAEEGPPA---APGSLRHS----------GPLGPHACPTAL-PEPQVISAMSSQVVGIEP 107

Qy         96 SHIYAEPQKSPWCEARSLEHTIPVNREFITKRKVSGNRCA--------SPVTGPG------ 141
                :: |||       :  |   | |     |    :  ||             |||
Db        108 LYIKAEPASPDSPKGSSETETEP--PVALAPGPAPTRCLPGHKEEEDGEGAGPGEQGGGK 165

Qy        142     SEEQAHCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRR 198
                  |      | || | |||||||| ||| ||||||||:||| :| |||:|: | | ||
Db        166 LVISSIPKPICLVCGDVASGYHYGVASCEACKAFFKRTIQGSIEYSCPASNECPITKRRR 225

Qy        199 KSCQACRLPKCYEVGMVKGGSRREPC-GYRLVPQQRSADFQL-----HCAGKAKPSGGHA 252
              |:|||||  ||  |||:| | | :|  | |   ::|   : |          ||    :||
Db        226 KACQACRFMKCLEVGMLKEGVRLDRVRGGRQKYKRRPDVDPLPPPCPEPAGPLAVAGG-- 283

Qy        253 PRVRELLLDALSPEQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLIKLADKELVHMIS 311
              ||      ::|      ||  ||  ||  :  :  |:|      : :|  | |:|:|  ||
Db        284 PRKTAAPVNA-----LVSHLLVVEPEKLYAMPDPAGPDGHLPAVATLCDLFDREIVVTIS 338

Qy        312 WAKKIPGFVELSLEDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVE 371
              ||| ||||  ||| ||: :|:| |||||::|:  ||:    :| || ||||| :||
Db        339 WAKSIPGFSSLSISDQMSVLQSVWMEVIVLGVAQRSLPLQDELAFAEDLVLD-EEGARAA 397

Qy        372 GILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLN 431
              |: |:   ||    | : |:|: :||: :||: | ||    :    :    |  || |
Db        398 GLGFLGAALIQLVRRIQALRLEREEYVLLKALALANSDSVHIEDFPRLWSSCEKLLH--- 454
```

```
Qy          432 AVIDALV WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 496
                :|||: :   ::|     :  |   ||: |  :|  : | : |   :| :  ||:: |
Db          455   EALLEYEAGRAGPGGGAERRRAGRLLLTLPLLRQTAGKVLAHFYGVKLEGKVPMHKL 511

Qy          491 LLEMLNA 497
                |||| |
Db          512 FLEMLEA 518


RESULT 17
A43781
retinoid X receptor-gamma - chicken
N;Alternate names: retinoic acid receptor X-alpha homolog
C;Species: Gallus gallus (chicken)
C;Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 02-Aug 2002
C;Accession: A43781; S46478
R;Rowe, A.; Eager, N.S.C.; Brickell, P.M.
Development 111, 771-778, 1991
A;Title: A member of the RXR nuclear receptor family is expressed in neural crest derive
A;Reference number: A43781; MUID:91347905; PMID:1652422
A;Accession: A43781
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-467 <ROW>
A;Cross-references: GB:X58997; NID:q63495; PIDN:CAA41743.1; PID:q63496
R;Seleiro, E.A.P.; Darling, D.; Brickell, P.M.
Biochem. J. 301, 283-288, 1994
A;Title: The chicken retinoid-X-receptor-gamma gene gives rise to two distinct species o
A;Reference number: S46478; MUID:94311845; PMID:8037682
A;Accession: S46478
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-17 <SEL>
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: DNA binding; transcription regulation; zinc finger
F;141-383/Domain: erbA transforming protein homology <EPBA>

  Query Match             16.8%;  Score 471;  DB 2;  Length 467;
  Best Local Similarity   28.5%;  Pred. No. 5.8e-29;
  Matches  146;  Conservative   72;  Mismatches  218;  Indels   76;  Gaps   15;

Qy            5 NSPSSLNSPSSYN-----CSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSI 58
                | |   |: :          | |: |    |:    | || || |      :|   : ::
Db            4 NYPHFIKEPAGPGNSPVHASSTSVSPSSSLSV--GSTVDGHHNY----LEAPTNASPAL 56

Qy           59 PSNVTNLEGGPGRQTTSP-NVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTL 117
                || : |  | |     ||  |:  : |          ||  |    :|      :
Db           57 PSPM NTIGSPVNALGSPYRVIASSIG------  -  SHPVALSSSAP--  GMNEVI 99

Qy          118 PVNRETLKRKVSGNRCASPV------     -  TGPGSKRDAHFCAVCSDYASGYHYGVW 164
                :  :   || :    |:              | |||    | ||:| | :|| ||||:
Db          100 HSPQPNVLNNVSSSEDIKPLPGLPGIGNMNYPSTSPGSLA-KHICAICGDRSSGKHYGVY 158

Qy          165 SCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGYRLVRRQRSADEQLHCAG 244
                |||||| |||:|:    | |    | ||| :|  || || :||  :|| :   :  ||
Db          159 SCEGCKGFFKRTIRKDLIYTCPDNKDCLIDKRQRNRCQYCRYQKCLAMGMKREAVQEERQ 219

Qy          225 GYRLVRRQRSADEQLHCAGKAKRSGCHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISR 284
                |     |:|| :|          |||      | :|:|          |  ||     |
Db          219 G--  SRERSENE---  AESTSGGSEDMPVERILEA---------ELAVEPKTEAYSD 259

Qy          285 PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLM 344
                :   :    : ::   |||:|  :: |||:|| | :|:| ||| || : |:|:
Db          260 VNTESSTNDPVTNICHAADKQLFTLVEWAKRIPHFSDLTLEDQVILLRAGWNELLIASFS 319

Qy          345 WRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAM 403
                 ||:     :: |  | : |       |:  ||| :|    |: ::::::   |  |:::|:
Db          320 HRSVSVQDGILLATGLHVHRGSAHSA GVGSIFDRVLTELVSKMKDMQMDKSELGCLRAI 378

Qy          404 ILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLS 463
                :| |    | : ::      |:   | | |                :|  | | ||: |
Db          379 VLFNPDAKGLSSPSEVESLREKVYATLEAYTKQ---------KYPEQPGRFAKLLLRLP 428

Qy          464 HVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                  :|    | :|||   |     |:   |:|||
Db          429 AIPSPIELCCLEHLFFFKLIGDTPIDTFLMEML 460


RESULT 18
A39596
progesterone receptor B form - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20 Mar 1992 #sequence_revision 20 Mar 1992 #text_change 20-Aug 1999
C;Accession: A39596; I49111
R;Schott, D.R.; Shyamala, G.; Schneider, W.; Parry, G.
Biochemistry 30, 7014-7020, 1991
A;Title: Molecular cloning, sequence analyses, and expression of complementary DNA en
A;Reference number: A39596; MUID:91299759; PMID:2069958
A;Accession: A39596
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-923 <SCH>
A;Cross-references: GB:M68915; GB:J05333; NID:g200471; PIDN:AAA39971.1; PID:g200472
R;Haqihara, K.; Wu-Peng, X.S.; Funabashi, T.; Kato, J.; Pfaff, D.W.
Biochem. Biophys. Res. Commun. 205, 1093-1101, 1994
A;Title: Nucleic acid sequence and DNase hypersensitive sites of the 5' region of the
A;Reference number: I49111; MUID:95100931; PMID:7802637
A;Accession: I49111
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-9 <HAG>
A;Cross-references: EMBL:U12644; NID:g639916; PIDN:AAA66067.1; PID:g639917
C;Superfamily: progesterone receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation;
F;555-819/Domain: erbA transforming protein homology <ERBA>
F;557-577/Region: zinc finger
F;593-617/Region: zinc finger

  Query Match             16.7%;  Score 469.5;  DB 2;  Length 923;
  Best Local Similarity   26.5%;  Pred. No. 1.8e-28;
  Matches  136;  Conservative   81;  Mismatches  162;  Indels  135;  Gaps   18;

Qy            6 SPSSLNSPSSYNC--SQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVT 63
                :|     |:: :|   :  ||   |:   |: |             |  ||:  |:
Db          470 APLPCKIPPAAASCLLPRDSLPAAPGTAAAPAIYQPLGLNGLPQLGYQAAVLKDSL----- 524

Qy           64 NLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRET 123
                          | |  |    :| |          :|  :||                 ::
Db          525 -----------PQVYPPYLNYLRP----------DSEASQSP-----------QYGFDS 551

Qy          124 LKRKVSGNRCASPVIGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDY 183
                | :|:                    | :| | ||| |||| :|  || ||||:::| ::|
Db          552 LPQKI-------------------CLICGDEASGCHYGVLTCGSCKVFFKRAMEGQHNY 591

Qy          184 ICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGYRLVRRQRSADEQLHCAG 243
                :|   | | :|| ||||:| |||||||  : |||  | : ::   :  ||  |: |      |
Db          592 LCAGRNDCIVDKIRRKNCPACRLRKCCQAGMVLGGRKFKK  FNKVRVMRTLD------G 643

Qy          244 KAKRSGCHAPPVRELLIDALSPEQ-------LVLTLLEAEPPHVLISPPSA-PFIEAS 293
                |        |   : |   : || |       |:  |:  ||  :     : | :|
Db          644 VALPQSVGLPNESQALSQRITFSPNQEIQLVPPLINLLMSIEPDVIYAGHDNTKPDTSSS 703

Qy          294 MMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG- 352
                :: || :| :::|: :: |:| :|||  | : || : |:::  || ::: || |||  |
Db          704 LLTSLNQLGERQLLSVVKWSKSLPGFRNLHIDDQITLIQYSWMSLMVFGLGWRSYKHVSG 763

Qy          353 -KLIFAPDLVLDRDEGKCVEGILEIFDMLLA---TTSRFRELKLQHKEYLCVKAMILLNS 408
                 | |||||:|:    |     |  : : |        | :|:: |:|:||:| ::|||:
Db          764 QMLYFAPDLILNEQRMK----ELSFYSLCLTMWQIPQEFVKLQVTHEEFLCMKVLLLLNT 819

Qy          409 SMYPLVIATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 468
                                                |   |: || |   : :     |:::
Db          820 --- ------- ----   -   -      IPLEGLRSQSQIEEMKS     SYIKE- 841

Qy          469 SNKGMEHLLNMKCKNVVPV--- YDL--LLEMLN 496
                    :  :: | |||     | |  ||: |:
```

```
Db        842 ----LIKAIGLRQKGVVPTSQRFYQLTKLLDSLH 871


RESULT 19
I53280
progesterone receptor B form - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 20-Aug-1999
C;Accession: I53280; B23733; A49574; A23733
R;Park-Sarge, O.K.; Mayo, K.E.
Endocrinology 134, 709-718, 1994
A;Title: Regulation of the progesterone receptor gene by gonadotropins and cyclic adenos
A;Reference number: I53280; MUID:94130817; PMID:8299566
A;Accession: I53280
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-923 <PAR1>
A;Cross-references: GB:L16922; NID:g463282; PIDN:AAA19916.1; PID:g463283
R;Park, O.K.; Mayo, K.E.
Mol. Endocrinol. 5, 967-978, 1991
A;Title: Transient expression of progesterone receptor messenger RNA in ovarian granulos
A;Reference number: A23733; MUID:92049379; PMID:1840636
A;Accession: B23733
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr
A;Molecule type: mRNA
A;Residues: 555-589,'W',591-621;742-910 <PAR2>
A;Cross-references: GB:S64044
R;Kraus, W.L.; Montano, M.M.; Katzenellenbogen, B.S.
Mol. Endocrinol. 7, 1603-1616, 1993
A;Title: Cloning of the rat progesterone receptor gene 5' region and identification of t
A;Reference number: A49574; MUID:94195318; PMID:8145766
A;Accession: A49574
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-16 <KRA>
A;Cross-references: GB:S69361; NID:g546178
C;Superfamily: progesterone receptor; erbA transforming protein homology
C;Keywords: DNA binding, nucleus, steroid hormone receptor, zinc finger
F;555-819/Domain: erbA transforming protein homology <ERBA>
F;557-577/Region: zinc finger
F;593-617/Region: zinc finger

  Query Match             16.6%;  Score 465;  DB 2;  Length 923;
  Best Local Similarity   33.4%;  Pred. No. 4.1e-28;
  Matches  110;  Conservative   62;  Mismatches  121;  Indels   36;  Gaps   11;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | :| | ||| |||| :|  || |||||:::| ::|:|   | | :|| |||:| ||||||
Db        557 CLICGDEASGCHYGVLTCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRK 616

Qy        209 CYEVGMVKCGSPRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDAL--SPE 266
              | : |||  | : ::   :  ||  |: |      | |        |   : |   :  ||
Db        617 CCQAGMVLGGRKFKK--FNKVRVMRALD------GVALPQPVAFPNESQTLGQKIFSPN 668

Qy        267 Q-------LVLTLLEAEPPHVLISRPSA-PFTEASMMMSLTKLADKELVHMISWAKKIPG 318
              |       |:  |:  ||  |     : | | :|:: || :| ---|- -- |-| -||
Db        669 QEIQLVPPLINLLMSIEPDVVYAGHDNTKPDTSSSLLTSLNQLGERQLLSVVKWSKSLPG 728

Qy        319 FVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGKCVEGIIEI 376
              |  | : ||: |::  ||  ::: || |||  |     | ||||||:|:    |     |
Db        729 PRNIHIDDQITLIQYSWMSLMVFGLGWRSYKHVSGQMLYFAPDLILNEQRMK----ELSF 784

Qy        377 FDMLLA---TTSPFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAV 433
              : : |        | :|:: |:|:|||:| ::|||:   ||    :   |  :   : ::
Db        785 YSLCLTMWQIPQEFVKLQVTHEEFLCMKVLLLLNT--IPL----RGLRSQSQFEEMRSSY 838

Qy        434 TDALVWVIA---KSGISSQQQSMRLANLL 459
                    |:  |    |  : | |:  :|  ||
Db        839 IRELIKAIGLRQKGVVPSSQRFYQLTKLL 867


RESULT 20
A34418
H-2 region II binding protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 15-Jun-1990 #sequence_revision 09-Oct-1992 #text_change 02-Aug-2002
C;Accession: A34418
R;Hamada, K.; Gleason, S.L.; Levi, B.Z.; Hirschfeld, S.; Appella, E.; Ozato, K.
Proc. Natl. Acad. Sci. U.S.A. 86, 8289-8293, 1989
A;Title: H-2RIIBP, a member of the nuclear hormone receptor superfamily that binds to
A;Reference number: A34418; MUID:90046782; PMID:2554307
A;Accession: A34418
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-446 <HAM>
A;Cross-references: GB:M26804; NID:g193740; PIDN:AAA37772.1; PID:g487184
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: DNA binding; transcription regulation; zinc finger
F;116-362/Domain: erbA transforming protein homology <ERBA>

  Query Match             16.4%;  Score 459;  DB 2;  Length 446;
  Best Local Similarity   29.9%;  Pred. No. 4.8e-28;
  Matches  155;  Conservative   65;  Mismatches  191;  Indels  108;  Gaps   21;

Qy          7 PSSLNSPSSYNCSQSILPLE-HGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVTNL 65
              | | :| |    || | |    |    ||:         | |    |  :    |  :::
Db          4 PSSPDSASPNFTSQCTFYSSFPGFPIISA      PTFPM --PPPPGSPTTV ISSS 52

Qy         66 PGSPGSQTTSPNVTWPTPSGLSFIVVHSQLSHIVAFPQKSPWTFAKSIPHTIPVN----R 121
              | ||    :|    ||                 :: |  ||    :     :||
Db         53 MGSPGLPPPAP------PG--------------FSGPVSSPQTNSTV---SLPGGGSGPP 89

Qy        122 EILKEKVSGNR    CASPVTGGSSERDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQ 178
              | :|  | | |    |  |  |||: :     ||:| | :|| ||||:|||||| ||||:|:
Db         90 PDVKPPVIGVRGLHCPPPPGGPGAGK--PLCAICGDRSSGCHYGVYSCDGCKGFFKRTIR 147

Qy        179 GHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQ 238
                | |     ||:|| :|  || || :||   || :   : |        ||| |:
Db        148 KDLIYSCRLNKDCIVDKRQRNRCQYCRYQKCLAIGMKREAVQEE------ --RQRGKDKD 199

Qy        239 LHCAGKAKRSGCHAPRVRELLLDALSPEQL-VLTLLEAEPPHVLI- ------- ----- 282
              |    :||              :|||: |  :||||   : :
Db        200 ----GDGDGAGG-------------APEEMPVDRILEAE-- LAVEQKSDQGVEGPGATG 239

Qy        283 ---SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVL 339
                 | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : | |:|
Db        240 GGGSSPNDFVT       NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGWNELL 293

Qy        340 MMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGIIEIFDMLLA-TTSRFRELKLQHKEYL 398
              :     ||||    :: | | : |:       |:  ||| :|    |: |::::   |
Db        294 IASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDKTELG 352

Qy        399 CVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL 458
              |:::|:|| |     |         |:     :|     | :    |     ||   | | |
Db        353 CLRAIILFNPDAKGL--------SNPGEVEILREKVYASLETYCKQKYPEQQG--RFAKL 402

Qy        459 LMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
              || |  :|    | :|||   |     |:   |:|||
Db        403 LLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 441


RESULT 21
A41651
retinoic acid receptor corequlator - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 02-Aug-2002
C;Accession: A41651
R;Yu, V.C.; Delsert, C.; Andersen, B.; Holloway, J.M.; Devary, O.V.; Naar, A.M.; Ki
Cell 67, 1251-1266, 1991
A;Title: RXRbeta: a coregulator that enhances binding of retinoic acid, thyroid hormo
A;Reference number: A41651; MUID:92103690; PMID:1662118
A;Accession: A41651
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-451 <YUA>
A;Cross-references: GB:M81766; NID:g206602; PIDN:AAA42025.1; PID:g206603
```

```
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: zinc finger
F;121-367/Domain: erbA transforming protein homology <ERBA>

  Query Match             16.4%;  Score 459;  DB 2;  Length 451;
  Best Local Similarity   29.2%;  Pred. No. 4.8e-28;
  Matches  143;  Conservative   62,  Mismatches  186,  Indels   98,  Gaps   16,

Qy          50 SPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPL-VVHRQLSHLYAEPQKSPWC 108
               ||  ::  ||    :   ||    ::|    | |   ||  |:   :        |   |
Db          15 SPNPLSQGIP  -PSSPPGPPHTPSAPPPPMPPPPLGSPFPVISSSMGSPGLPPPAPPGF 71

Qy         109 EARSLEHTLPVNRETLERKVS- --- --  -- -------GNRCASPVTGPGSKRDAHF 148
                        ||:   :  ||                    |  |  |  |||: :
Db          72 SG      PVSSPQINSTVSLPGGGSGPPEDVKPPVLGVRGLHCPPPPGGPGAGK  RL 122

Qy         149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
               |:| | :|| |||||:|||||| ||||:|:    | |     ||:|| :|  || || :|
Db         123 CAICGDRSSGKHYGVYSCEGCKGFFKRTIRKDLTYSCRDNKDCTVDKRQRNRCQYCRYQK 182

Qy         209 CYEVGMVKCGSKRERCGYRLVKRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQL 268
               |   || :   : |          |||  |:     |    :||          :||::
Db         183 CLATGMKREAVQEE  -       RQRGKDKD-  - GDGDGAGG------- ---  APEEM 217

Qy         269  VLILLEAEPPHVLI -         - - -----SRPSAPFTEASMMMSLTKLADKELVHM 309
                |  :||||   :  :                | |: | |      :: : |||:|  :
Db         218 PVDRILEAE-  -LAVEQKSDQGVEGPGATGGGGSSPNDPVT------NICQAADKQLFTL 268

Qy         310 ISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDREEGKC 369
               : |||:|| |   | | ||| || : | |:|:     ||||    :: |   | : |:
Db         269 VEWAKRIPHFSSLPLDDQVILLRAGWNELLIASFSHRSIDVRDGILLATGLHVHRNSAHS 328

Qy         370 VEGILEIFDMLLA TTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAH 428
                 |:  ||| :|    |: |::::   |  |::|:||| |     |         |:
Db         329 A GVGAIFDRVLTELVSKMRDMRMDKTELGCLRAIILFNPDAKGL--------SNPGEVE 379

Qy         429 LLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVY 488
               :|     | :    |     ||   | | ||: |  :|    | :|||   |     |:
Db         380 ILREKVYASLETYCKQKYPEQQG--RFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPID 437

Qy         489 DLLLEMLNA 497
                 |:||| |
Db         438 TFLMEMLEA 446


RESULT 22
S47781
retinoid X receptor beta - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 02-Aug-2002
C;Accession: S47781; S39753; S22647; A41727; E41727; A44270
R;Fleischhauer, K.; McBride, O.W.; DiSanto, J.P.; Ozato, K.; Yang, S.Y.
Hum. Genet. 90, 505-510, 1993
A;Title: Cloning and chromosome mapping of human retinoid X receptor beta: selective ami
A;Reference number: S47781; MUID:94154716; PMID:8381486
A;Accession: S47781
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-533 <FLE1>
A;Cross-references: EMBL:X63522
R;Fleischhauer, K.
submitted to the EMBL Data Library, December 1991
A;Reference number: S39753
A;Accession: S39753
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-117,'A',119-533 <FLE2>
A;Cross-references: EMBL:X63522; NID:g40447; PIDN:CAA45087.1; PID:g40448
R;Fleischhauer, K.; Park, J.H.; DiSanto, J.P.; Marks, M.; Ozato, K.; Yang, S.Y.
Nucleic Acids Res. 20, 1801, 1992
A;Title: Isolation of a full length cDNA clone encoding a N-terminally variant form of t
A;Reference number: S22647; MUID:92253386; PMID:1315958
A;Accession: S22647
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-90 <FLE3>
A;Cross-references: EMBL:X63522
R;Leid, M.; Kastner, P.; Lyons, R.; Nakshatri, H.; Saunders, M.; Zacharewski, T.; Che
Cell 68, 377-395, 1992
A;Title: Purification, cloning, and RXR identity of the HeLa cell factor with which R
A;Reference number: A41727; MUID:92127595; PMID:1310259
A;Accession: A41727
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-15,'E',17,'SV',20,35-36,'GAKECIVGS',46,'T',48,'L',50,'GSRS',55-56,'GGGG
A;Note: this sequence has been revised in reference A44270
A;Note: sequence extracted from NCBI backbone (NCBIP:77545)
A;Accession: E41727
A;Molecule type: protein
A;Residues: 272-277;304-311;393-413;420-426;443-452;494-497 <LE2>
R;Leid, M.; Kastner, P.; Lyons, R.; Nakshatri, H.; Saunders, M.; Zacharewski, T.; Che
Cell 5, 886a, 1992
A;Reference number: A44270
A;Contents: erratum
A;Accession: A44270
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-76 <LE3>
A;Cross-references: GB:M84820
C;Genetics:
A;Gene: GDB:RXRB
A;Cross-references: GDB:128801; OMIM:180246
A;Map position: 6p21.3-6p21.3
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: zinc finger
F;203-449/Domain: erbA transforming protein homology <ERBA>

  Query Match             16.3%;  Score 458.5;  DB 2;  Length 533;
  Best Local Similarity   30.2%;  Pred. No. 6.6e-28;
  Matches  140;  Conservative   63;  Mismatches  163;  Indels   97;  Gaps   17;

Qy          62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
               :::  | ||    :|     ||              :: |  ||   :     :||
Db         136 ISSSMGSPGLPPPAP-----PG-------------PSGPVSSPQINSIV---SLPGGG 172

Qy         121 ---RETLERKVSCNR---CASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                  | :|  | | |   |  |  |||: :    ||:| | :|| ||||:|||||| |||
Db         173 SGPPEDVKPPVLGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFK 230

Qy         175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRS 234
               |:|:    | |     ||:|| :|  || || :||   || :   : |          |||
Db         231 RTIRKDLTYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEE--------RQRG 282

Qy         235 ADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQL VLTLLEAEPPHVLI             282
                |:      |  : :||             :|||: |  :||||   : :
Db         283 KDKD----GDGEGAGG-------------APEEMPVDRILEAE---LAVEQKSDQGVEGP 322

Qy         283 -------SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
                      | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : |
Db         323 GGTGGSGSSPNDPVT- --- -NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 376

Qy         336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDREGKCVEGILEIFDMLLA-TTSRFRELKLQH 394
                 |:|:     ||||    :: |  | : |:        |:  ||| :|    |:  |::::
Db         377 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA GVGAIFDRVLTELVSKMRDMRMDK 435

Qy         395 KEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR 454
               |  |::|:|| |     |        |:      :|     |  :    |     ||   |
Db         436 TELGCLRAIILFNPDAKGL--------SNPSEVEVLREKVYASLETYCKQKYPEQQG--R 485

Qy         455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
                | ||: |  :|    | :|||   |     |:   |:||| |
Db         486 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 528


RESULT 23
S60586
```

```
glucocorticoid receptor - rainbow trout
C;Species: Oncorhynchus mykiss (rainbow trout)
C;Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 20 Aug-1999
C;Accession: S60586; S70470
R;Ducouret, B.; Tujague, M.; Ashraf, J.; Mouchel, N.; Servel, N.; Valotaire, Y.; Pakdel
Endocrinology 136, 3774-3783, 1995
A;Title: Cloning of a teleost fish glucocorticoid receptor shows that it contains a deox
A;Reference number: S60586; MUID:96377204; PMID:7649084
A;Accession: S60586
A;Molecule type: mRNA
A;Residues: 1-758 <DUC>
A;Cross-references: EMBL:Z54210; NID:g995676; PIDN:CAA90937.1; PID:g995677
R;Takeo, J.; Hata, J.; Segawa, C.; Toyohara, H.; Yamashita, S.
FEBS Lett. 389, 244-248, 1996
A;Title: Fish glucocorticoid receptor with splicing variants in the DNA binding domain.
A;Reference number: S70470; MUID:96335602; PMID:8766708
A;Accession: S70470
A;Molecule type: mRNA
A;Residues: 1-89,'L',91-246,'P',248-281,'Y',283-303,'T',305-351,'RLI',355,'S',357-444,'S
A;Experimental source: pituitary
C;Genetics:
A;Gene: GR
C;Superfamily: glucocorticoid receptor; erbA transforming protein homology
C;Keywords: alternative splicing; DNA binding; steroid binding; transcription regulation
F;59-66/Region: glutamine-rich
F;385-649/Domain: erbA transforming protein homology <ERBA>
F;387-407/Region: zinc finger CCCC motif
F;432-456/Region: zinc finger CCCC motif
F;703-752/Domain: steroid binding #status predicted <STR>

  Query Match             16.3%;  Score 457.5;  DB 2;  Length 758;
  Best Local Similarity   29.4%;  Pred. No. 1.2e-27;
  Matches  146;  Conservative   71;  Mismatches  161;  Indels  119;  Gaps   19;

Qy          8 SSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSH--  -      HEYPAMTFYSPAVMNYSI 58
              ||||: ||:|| :         |||  ||    ::          : |    | |      :
Db        286 SSLDLPSTHNSA -       -GSISGPSYPYGANTSTAVSIQQDQKPVFGLYPP------I 331

Qy         59 PSNVTNLEGGPGRQTTS--PNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEAKSLEHT 116
              ||   :  | | | |   :  :|                 :: |:  |
Db        332 PSVSDSWNRGNGYATGSGMSSSSFPVG--------------FSSPKARP---------- 366

Qy        117 LPVNRETLKPKVSGNPCASPV--TGPGSKFDAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                  : ||:  :::|   :||  |     | |||| ||| |||| :|  || |||
Db        367 ----------EASGSASSAPAKPSGPTHK----ICLVCSDEASGCHYGVITCGSCKVFFK 412

Qy        175 RSIQG--    -----HNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGRRC 224
              |:::|          || |:|   | | ||| |||:| ||| ||| : ||    :|: :
Db        413 RAVEGWPARQNTDQQHN-YICAGRNDCIIDKIRRKNCPACRFRKCLQAGM-NLEAREKK 470

Qy        225 GYRLVRRQRSADPQSHCACKAKFRGGHAPEVREILLDALSPEQLV   :TLLEAEPPHVI 282
               ||  :| : :                    | |   |   | |||    |:|||:|  | :
Db        471 LIRLKGQQTTME----------PNPPPPDERACALIPKSMPQLVPTMLSLLKAIEPEAI 519

Qy        283 IS--RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVL 340
               |    : | |   :| :| :|  :::|  : ||| :|||  | | ||: ||:  |: ::
Db        520 YSGYDSTIPDTSTRIMTTLNRLGGQQVVSAVKWAKSLPGFRNLHLDDQMILLQCSWLFLM 579

Qy        341 MMGLMWRSID  HPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEY 397
               || |||    : | | |||||:: ||   :  : :  : :|  :: |  |:: : ||
Db        580 SFGLGWRSYQQCNGGMLCFAPDLVIN-DERMKLPYMTDQCFQMIKISTEFVRLQVSYEEY 638

Qy        398 LCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTD--------ALVWVIAKSGISSQ 449
               ||:| ::|||:       |  :|    |:        || |          |   | |  :|
Db        639 LCMKVLLLLS-------IVPKDGLKSQ-------AVFDEIRMTYIKELGKAIVKREENSS 684

Qy        450 QQSMRIANLLMLLSHVR 466
               |   |   |  ||  ::
Db        685 QNWQRFYQLTKLLDSMQ 701


RESULT 24
D41727
```

```
retinoid X receptor beta - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Aug-2002
C;Accession: D41727
R;Leid, M.; Kastner, P.; Lyons, R.; Nakshatri, H.; Saunders, M.; Zacharewski, T.; Che
Cell 68, 377-395, 1992
A;Title: Purification, cloning, and RXR identity of the HeLa cell factor with which R
A;Reference number: A41727; MUID:92127595; PMID:1310259
A;Accession: D41727
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-448 <LEI>
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:77547)
C;Genetics:
A;Start codon: CTG
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: zinc finger
F;118-364/Domain: erbA transforming protein homology <ERBA>

  Query Match             16.3%;  Score 456;  DB 2;  Length 448;
  Best Local Similarity   30.2%;  Pred. No. 8.2e-28;
  Matches  140;  Conservative   62;  Mismatches  164;  Indels   97;  Gaps   17;

Qy         62 VINLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEAKSLEHTLPVN- 120
              :::  | ||    :|      ||              :: |  ||   :      :||
Db         51 ISSSMGSPGLPPPAP------PG--------------FSGPVSSPQINSTV---SLPGGG 87

Qy        121    EELIEEEVVGNE - -CASPVIGPGSEEEDAHFCAVCSLYASGYHYGVWSCEGCKAFFK 174
                 | :|  | | |   |  | |||: :    ||:| | :|| ||||:|||||| |||
Db         88 SGFPEDVKPPFVLGGVEGLHCPPPPGGPGAGK--RICAICGDPSSGKHYGVYSCEGCKGFFK 145

Qy        175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGRRIVRRQRS 234
              |:|:    | |     |:|| :|  || || :||   || :   : |         |||
Db        146 RTIRKDLIYSCRDNKDCIVDKRQRNRCQYCRYQKCLATGMKREAVQEE--------RQRG 197

Qy        235 ADPQSHCACKAKFRGCHAPRVRELLLDALSPEQL-VLTLLEAEPPHVI------------ 282
               |:     |    :||              :|||:: |  :||||    : :
Db        198 KDKD----GDGDGAGG-------------APEEMPVDPILEAE---IAVFQKSDQGVEGP 237

Qy        283 -------SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
                    | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : |
Db        238 GATGGGGSSPNDPVI -     NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 291

Qy        336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-ITSRFRELKLQH 394
               |:|:    ||||     :: | | : |:        |:  ||| :|    |: |::::
Db        292 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDK 350

Qy        395 KEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR 454
               |  |::|:|| |     |         |:     :|     | :    |     ||   |
Db        351 IFLGCLRAIILFNPDAKGL -----SNPGEVEILREKVYASLETYCKQKYPEQGR--R 400

Qy        455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
               | ||: |  :|    | :|||   |     |:   |:||| |
Db        401 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 443


RESULT 25
I84718
RXR-beta1 isoform - mouse
C;Species: Mus musculus (house mouse)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 20-Aug-1999
C;Accession: I84718
R;Nagata, T., Kanno, Y., Ozato, K., Taketo, M.
Gene 142, 183-189, 1994
A;Title: The mouse Rxrb gene encoding RXR beta: genomic organization and two mRNA iso
A;Reference number: I48752; MUID:94252565; PMID:8194750
A;Accession: I84718
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-520 <RES>
A;Cross-references: GB:D21831; NID:g497828; PIDN:BAA04858.1; PID:g987669
C;Genetics:
```

```
A;Gene: Rxrb
A;Introns: 76/3; 148/3; 261/1; 264/1; 402/1; 466/2; 492/1; 527/2
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
F;190-436/Domain: erbA transforming protein homology <ERBA>

  Query Match              16.4%;  Score 456;  DB 2;  Length 520;
  Best Local Similarity    30.2%;  Pred. No. 1e-27;
  Matches  140;  Conservative   62;  Mismatches  164;  Indels   97;  Gaps   17;

Qy          62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN- 120
               :::  | ||    :|      ||               :: |  ||   :     :|| 
Db         123 ISSSMGSPGLPPPAP---  --PG-------  -------FSGPVSSPQINSTV---SLPGGG 159

Qy         121     RETLKRKVSGNR   CASPVTGPGSKRGAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                 | :|  | | |    |  |  ||| : :    ||:| | :|| |||||||||| |||
Db         160 SGPPEDVKPPVLGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFK 217

Qy         175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVRRQRS 234
               |:|:    | |     ||:|| :|  || || :||   || :   : |         |||
Db         218 RTIRKDLIYSCRDNKDCTVDKRQRNRCQYCRYQKCLAIGMKREAVQEE------- RQRG 269

Qy         235 ADEQLHCAGKAKRSGGHAPKVRELLLDALSPEQL-VLTLLEAEPPHVLI------- -    282
               |:     |   :||          :|||: |  :||||    : :
Db         270 KDKD    GDGDGAGG-------------APEEMPVDRILEAE---LAVEQKSDQGVEGP 309

Qy         283         SRPSAPETPFASMMMSLTKLADKELVHMISWAKRIPGFVELSLFDQVRLLESCW 335
                     | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : |
Db         310 GAIGGGGSSPNDPVT-----NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 364

Qy         336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQH 394
                |:|:    ||||    :: |  | : |:       |:  ||| :|    |: |:::: 
Db         364 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDK 422

Qy         395 KEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR 454
               |  |::|:|| |     |          |:     :|     | :    |     ||   |
Db         423 TELGCLRAIILFNPDAKGL        -SNPGEVEILREKVYASLETYCKQKYPEQQG--R 472

Qy         455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
               | ||: |  :|    | :|||   |     |:   |:||| |
Db         473 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 515


RESULT 26
S26669
retinoic acid receptor X beta - mouse
N;Alternate names: RXR-beta2 isoform
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C;Accession: S26669; I63160
R;Mangelsdorf, D.J.; Borgmeyer, U.; Heyman, R.A.; Zhou, J.Y.; Ong, E.S.; Oro, A.E.; Kaki
Genes Dev. 6, 329-344, 1992
A;Title: Characterization of three RXR genes that mediate the action of 9-cis retinoic a
A;Reference number: S26668; MUID:92192447; PMID:1312497
A;Accession: S26669
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-410 <MAN>
A;Cross-references: EMBL:X66224; NID:g54023; PIDN:CAA46963.1; PID:g54024
R;Nagata, T.; Kanno, Y.; Ozato, K.; Taketo, M.
Gene 142, 183-189, 1994
A;Title: The mouse Rxrb gene encoding RXR beta: genomic organization and two mRNA isotor
A;Reference number: I48752; MUID:94252565; PMID:8194756
A;Accession: I63160
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-410 <RES>
A;Cross-references: GB:D21831; NID:g497828; PIDN:BAA04859.1; PID:g987670
C;Genetics:
A;Gene: Rxrb
A;Introns: 48/3; 91/1; 151/1; 252/1; 296/3; 327/1; 463/3
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: zinc finger
F;80-326/Domain: erbA transforming protein homology <ERBA>

  Query Match              16.2%;  Score 455.5;  DB 2;  Length 410;
  Best Local Similarity    30.2%;  Pred. No. 8e-28;
  Matches  140;  Conservative   62;  Mismatches  164;  Indels   97;  Gaps   17;

Qy          62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
               :::  | ||    :|      ||               :: |  ||   :     :|| 
Db          13 ISSSMGSPGLPPPAP---  --PG-------  -------FSGPVSSPQINSTV---SLPGGG 49

Qy         121 ---RETLKRKVSGNR---CASPVTGPGSKRGAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
                  | :|  | | |    |  |  ||| : :    ||:| | :|| |||||||||| |||
Db          50 SGPPEDVKPPVLGVRGLHCPPPPGGPGAGK--RLCAICGDRSSGKHYGVYSCEGCKGFFK 107

Qy         175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVRRQRS 234
               |:|:    | |     ||:|| :|  || || :||   || :   : |         |||
Db         108 RTIRKDLIYSCRDNKDCTVDKRQRNRCQYCRYQKCLAIGMKREAVQEE------ --RQRG 159

Qy         235 ADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQL VLTLLEAEPPHVLI---------- 282
               |:     |   :||          :|||: |  :||||    : :
Db         160 KDKD----GDGDGAGG---- --------APEEMPVDRILEAE---LAVEQKSDQGVEGP 199

Qy         283 -------SRPSAPETPFASMMMSLTKLADKELVHMISWAKRIPGFVELSLFDQVRLLESCW 335
                     | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : |
Db         200 GAIGGGGSSPNDPVT------NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 253

Qy         336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA TTSRFRELKLQH 394
                |:|:    ||||    :: |  | : |:       |:  ||| :|    |: |:::: 
Db         254 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDK 312

Qy         395 KEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR 454
               |  |::|:|| |     |          |:     :|     | :    |     ||   |
Db         313 TELGCLRAIILFNPDAKGL--------SNPGEVEILREKVYASLETYCKQKYPEQQG--R 362

Qy         455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
               | ||: |  :|    | :|||   |     |:   |:||| |
Db         363 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 405


RESULT 27
JG0194
androgen receptor - Japanese eel
C;Species: Anguilla japonica (Japanese eel)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C;Accession: JG0194
R;Todo, T.; Ikeuchi, T.; Kobayashi, T.; Nagahama, Y.
Biochem. Biophys. Res. Commun. 254, 378-383, 1999
A;Title: Fish androgen receptor: cDNA cloning, steroid activation of transcription in
A;Reference number: JG0194; MUID:99119319; PMID:9918846
A;Accession: JG0194
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-848 <TOD>
C;Superfamily: unassigned erbA related proteins; erbA transforming protein homology
F;491-745/Domain: erbA transforming protein homology <ERB>

  Query Match              16.1%;  Score 453;  DB 2;  Length 848;
  Best Local Similarity    26.6%;  Pred. No. 3.2e-27;
  Matches  150;  Conservative  102;  Mismatches  218;  Indels   94;  Gaps   22;

Qy           1 MDLKNSPSSLNSPSSYNCS-   - QSTLPL -EHGSTYTPSSYVDSHHEYPAMTFYSPAV 58
               |: |:|    |   ::: |     : :||:   | |  : |  |:    :: : : || :
Db         299 MEHKDSEKCANMDGAHSTSVFSQFEQLLPVNASHYSQNV-SVRVEPQSDFSPTLYKSPGI 357

Qy          59 MNYSIPSNV---TNLEGGPGRQTTSP--------NVLWPTPGHLSPLVVHRQLSHLYAEP 102
                  :   ||     ::   |: |:          |    ||          |:  :  | :
Db         358 QKNAEKYNVQYDATFKSEDGKTTSEREWGFQYRYNFESCSTPSAPPHHCAHQNRAGPYNQF 417

Qy         103 QKSPW---------CEARSLEHTLPVN------RETLKRKVSGNRCASPVTGPG---SK 143
                 :|:          |  ||||  | |          :|| :: |:| : |         :
Db         418 FFNPFFYAFSCVVSEICYSFFMCFFNNLAEIPYSGSLKNFL-GDPLSGGYFCVSYRYEGE 476

Qy         144 RDAHF  ----  --  -CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCT 192
                  |:  |           | :| | ||| |||   :|  || ||||: :|   |:| : | ||
```

```
Db        477 PENVFPVFEFPPPQPTCLIGNEPASCDHKTATIGGMKVEFPKAAFPEMFYLVAFINLCI 536

Qy        193 IDKNPRKSCQACRLPKCYEVGMVKGGSRKERPGGYPLVPRQRSAPEQLHCAGKAKESGCHA 252
              ||| |||:| :|||:::|:  ||     : :: |       |  | |           |
Db        537 IDKLRRKNCPSCRLKKCFAAGMTLGARKLKKLG------QMKAREDT---------GQGQG 581

Qy        253 PRVRELLLDALSP--------FQLVITLIEAEPPHVLISRPP--SAPFTEASMMMSLTKLA 302
              |   ||   ::||         : : | :|||  | |: :       | : ||:: || :|
Db        582 PAEAEL---SVSPKYDLGFHTQSMFLNILEAIEPEVVNAGHDYGQPDSAASLLTSLNELG 638

Qy        303 DKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDH--PGKLIFAPDL 360
              ::::|| :: ||| :|||  | : ||: :::  || |::  | |||  .     | |||||
Db        639 PPQIVKVVKWAKGMPGFRELSLYDQMTVIQKHWMAVMVFALGWRSFKNVKSKMLYFAPDL 698

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              | :    : |  : |    :   :  |  |::  :|:|||:|||:| ::     :
Db        699 VFNEHRMQ-VSTMYEHCIRMKNFSQEFAMLQVSQEFLCMKALLLFSTIPVEGLKQQNFF 757

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQ-QQSMRLANLLMLLSHVRHASNKGM    EH 475
              |  |:    :| : | ||   :||  | : ||  || : |  :    |     :    ::
Db        758 DFIRR--SYINFI-DRIVSPPSFSSCSEFQQLTPLLDSLQPVLKELHQFTFDLFVQSQN 814

Qy        476 LLNMKCKNVVPVYDLLLEMLNAHV 499
              | |  |       ::: |::: ||
Db        815 LSNQVC- ---FPEMISELLSVHV 838


RESULT 28
A25923
progesterone receptor - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 20-Aug-1999
C;Accession: A25923
R;Loosfelt, H.; Atger, M.; Misrahi, M.; Guiochon-Mantel, A.; Meriel, C.; Logeat, F.; Ben
Proc. Natl. Acad. Sci. U.S.A. 83, 9045-9049, 1986
A;Title: Cloning and sequence analysis of rabbit progesterone-receptor complementary DNA
A;Reference number: A25923; MUID:87067449; PMID:3538016
A;Accession: A25923
A;Molecule type: mRNA
A;Residues: 1-930 <LOO>
A;Cross-references: GB:M14547; NID:g165631; PIDN:AAA31443.1; PID:g165632
C;Superfamily: progesterone receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation; Z
F;566-826/Domain: erbA transforming protein homology <ERBA>
F;568-588/Region: zinc finger
F;604-628/Region: zinc finger

  Query Match             16.1%;  Score 453;  DB 2;  Length 930;
  Best Local Similarity   31.6%;  Pred. No. 3.6e-27;
  Matches  104;  Conservative   66;  Mismatches  119;  Indels   40;  Gaps   10;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | :| | |||:|||| :|  || |||:::| :::|   | | :|| |||:| ||||||
Db        568 CLICGDEASGCHYGVLTCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRK 627

Qy        209 CYEVGMVKGGSRRFRGGYRLVRRQRSAD-----EQLHCAGKAKRSGGHAPPVPELLLDAL 263
              | : |||  | : ::  :  ||  |: |     : :    :::|
Db        628 CQQAGMVLGGRKFKK--FNKVRVMRALDAVALPQPVGIPNESQRI------------TF 672

Qy        264 SPEQ-------LVLTLLEAEPPHVLISRPSA-PFTEASMMMSLTKLADKELVHMISWAKK 315
              || |       |:  |:  ||  :     :  | | :|:: || :| :::|: :: |:|
Db        673 SPSQEIQLIPPLINLLMSIEPDVIYAGHDNTKPDTSSSLLTSLNQLGERQLLSVVKWSKS 732

Qy        316 IPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGKCVEGI 373
              :||| | : ||: |::  || ::: || ||| |     | ||||||:|:    |
Db        733 LPGFRNLHIDDQITLIQYSWMSLMVFGLGWRSYKHVSGQMLYFAPDLILNEQRMK-ESSF 791

Qy        374 LEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAV 433
              :   :      | :|:  :|:||:| ::|||   ||    :  |  :    : ::
Db        792 YSLCLTMWQIPQEFVKLQVSQEFFLCMKVLLLLNT  IPL    EGLESQTQFEEMRSSY 845

Qy        434 TDALVWVIA---KSGISSQQQSMRLANLL 459
                 |:  |    |  :|| |:  :|  ||
```

```
Db        21: IELLIEALLDAQMGVVSSQRPYQLIKLL 874


RESULT 29
S26670
retinoic acid receptor X-gamma - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C;Accession: S26670
R;Mangelsdorf, D.J.; Borgmeyer, U.; Heyman, R.A.; Zhou, J.Y.; Ong, E.S.; Oro, A.E.; K
Genes Dev. 6, 329-344, 1992
A;Title: Characterization of three RXR genes that mediate the action of 9-cis retinoi
A;Reference number: S26668; MUID:92192447; PMID:1312497
A;Accession: S26670
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-463 <MAN>
A;Cross-references: EMBL:X66225; NID:g54025; PIDN:CAA46964.1; PID:g54026
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: zinc finger
F;137-429/Domain: erbA transforming protein homology  ERBA-

  Query Match             16.1%;  Score 452;  DB 2;  Length 463;
  Best Local Similarity   27.9%;  Pred. No. 1.8e-27;
  Matches  141;  Conservative   72;  Mismatches  215;  Indels   78;  Gaps   15;

Qy          7 PSSLNSPSSYNCSQSILPLEHGSIYIPSSY VDSHHEYPAMTFYSPAVMNYSIISNVTNL 65
              |:       :  | |: |    |: :|:   :||| |      :|  :  |      |
Db         12 PTGPGGSPGHTGSISMSP----SVALPTGKPMDSHPSYTDPPVSAPRTL--SAVGTPLNA 65

Qy         66 EGCPCRQTTSPNVLWPTPCHLS--PLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRET 123
               | |  ||   :  |  | |:  | :      :| | |                  ::
Db         66 LGSPYRVIIS--AMGPPSGALAAPPGI------NLVAPPS---------------SQLN 101

Qy        124 IKPKVSGNPCASPV-------------TGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCK 170
              :  || :   |:              | |||   | ||:| | :|| ||||:||||||
Db        102 VVNSVSSSEDIKPLPGLPGIGNMNYPSTSPGS-LVKHICAICGDPSSGKHYGVYSCEGCK 160

Qy        171 AFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRFRGGYRLVR 230
              ||||:|:   | |    | ||| :|  || || :||  :|| :  : ||      |
Db        161 GFFKRTIRKKLIYTCPDNKDCLIDKRQRNRCQYCRYQKCLVMGMKREAVQEER-----QR 215

Qy        231 RQRSADEQLHCAGKAKRSGGHAPEVPELLTDALSPEQLVLTLLEAEPPHVLISRPSAPFT 290
              :  |: :  ||     |  |     |           |  ||       :  :
Db        216 SRERAESEAECA-----SSSHEDMPVERILEA---------ELAVEPKTESYGDMNVENS 261

Qy        291 EASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDH 350
                : ::   |||:|  :: |||:|| | :|:| ||| || : | |:|:     ||:
Db        262 TNDPVTNICHAADKQLFTLVEWAKRIPHFSDLTLEDQVILLRAGWNELLIASFSHRSVSV 321

Qy        351 PGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSS 409
              ::| |  | : |      |:  ||| :|    |: ::::    |  |:::|:| |
Db        322 QDGILLATGLHVHRSSAHS-RGVGSIFDRVLTELVSKMKDMQMDKSELGCLRAIVLFNPD 380

Qy        410 MYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHAS 469
              |   ::      |:  | | |              :|  | | ||: |  :|
Db        381 AKGLSNPSEVETLREKVYATLEAYTKQ----------KYPEQPGRFAKLLLRLPALRSIG 430

Qy        470 NKGMEHLLNMKCKNVVPVYDLLLEML 495
               | :|||   |     |:   |:|||
Db        431 LKCLEHLFFFKLIGDTPIDSFLMEML 456


RESULT 30
B41727
retinoid-X receptor-gamma - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1993 #text_change 02-Aug-2002
C;Accession: B41727; A40702; B40702
R;Leid, M.; Kastner, P.; Lyons, R.; Nakshatri, H.; Saunders, M.; Zacharewski, T.; Che
Cell 68, 377-395, 1992
A;Title: Purification, cloning, and RXR identity of the HeLa cell factor with which R
A;Reference number: A41727; MUID:92127595; PMID:1310259
```

```
A;Accession: B41727
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1 463 <LEI>
A;Cross-references: GB:M84819; NID:g200881; PIDN:AAA40082.1; PID:g200882
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:77550)
R;Lin, Q.; Linney, E.
Mol. Endocrinol. 7, 651-658, 1993
A;Title: The mouse retinoid X receptor gamma gene: genomic organization and evidence for
A;Reference number: A40702; MUID:93402774; PMID:8491126
A;Accession: A40702
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1 463 <LIU>
A;Cross-references: GB:S62948; NID:g386366; PIDN:AAB27244.1; PID:g386367
A;Experimental source: C57BL/10J, skeletal muscle
A;Note: sequence extracted from NCBI backbone (NCBIN:144225, NCBIP:144226)
A;Accession: B40702
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 124-463 <LI2>
A;Cross-references: GB:S62952; NID:g386368; PIDN:AAB27245.1; PID:g386369
A;Experimental source: C57BL/10J, skeletal muscle
A;Note: sequence extracted from NCBI backbone (NCBIN:144227, NCBIP:144228)
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: zinc finger
F;147-379/Domain: erbA transforming protein homology <ERBA>

  Query Match             16.1%;  Score 452;  DB 2;  Length 463;
  Best Local Similarity   27.9%;  Pred. No. 1.8e-27;
  Matches  141;  Conservative   72;  Mismatches  215;  Indels   78;  Gaps   15;

Qy          7 PSSLNSPSSYNCSQSILPLEHGSIYIPSSY-VDSHHEYPAMTFYSPAVMNYSIPSNVINL 65
              |:       :  | |: |    |: :|:   :||| |      :|  :  |       |
Db         12 PTGPGGSPGHTCSTSMSP----SVALPTCKPMDSHPSYTIPTPVSAPPTI--SAVCTPTNA 69

Qy         66 EGGPGRQTTSPNVLWPTPGHLS--PLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRET 123
               | | |  ||   : |  | |:  | :      :| | |                 ::
Db         66 LGSPYRVITS  AMGPPSGALAAPPGI-----NLVAPPS-------  -------SQLN 101

Qy        124 LKRKVGGNRCASPV           --  TGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCK 170
              :  || :   |:                | |||    | ||:| | :|| ||||:||||||
Db        102 VVNSVSSSSEDIKPLPGLPGIGNMNYPSTSPGS-LVKHICAICGDRSSGKHYGVYSCEGCK 160

Qy        171 AFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVR 230
              ||||:|:    | |     | ||| :|  || || :||  :|| :   : ||      |
Db        161 GFFKRTIRKDLIYTCRDNKDCLIDKRQRNRCQYCRYQKCLVMGMKREAVQEER-----QR 215

Qy        231 RQKSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFT 290
              :  |: :  ||     | |     | :|:|        |  ||          :  :
Db        216 SRFRAFSPAFCA-----SSSHFDMPVEPILEA---------ELAVEPKTESYGDMNVENS 261

Qy        291 EASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDH 350
              : ::   |||:|  :: |||:|| | :|:| ||| || : | |:|:     ||:
Db        262 TNDFVTNICHAADKQLFTLVEWAKRIPHFSDLTLEDQVILLRAGWNELLIASFSHRSVSV 321

Qy        351 PGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSS 409
              ::  |  | :|       |:  ||| :|    |: :::::   |  |::|::| |
Db        322 QDGILLATGLHVHRSSAHSA-GVGSIFDRVLTELVSKMKDMQMDKSELGCLRAIVLFNPD 380

Qy        410 MYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHAS 469
                |   ::      |:   | | |               :|  | | ||: |  :|
Db        381 AKGLSNPSEVETLREKVYATLEAYTKQ----------KYPEQPGRFAKLLLRLPALRSIG 430

Qy        470 NKGMEHLLNMKCKNVVPVYDLLLEML 495
              | :|||   |     |:   |:|||
Db        431 LKCLEHLFFFKLIGDTPIDSFLMEML 456


RESULT 41
ORHUP
progesterone receptor form B - human
```

```
N;Alternate names: hPR
N;Contains: progesterone receptor form A
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1987 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999
C;Accession: S09971, S12464, A03245
R;Kastner, P.; Krust, A.; Turcotte, B.; Stropp, U.; Tora, L.; Gronemeyer, H.; Chambon
EMBO J. 9, 1603-1614, 1990
A;Title: Two distinct estrogen regulated promoters generate transcripts encoding the
A;Reference number: S09971; MUID:90228361; PMID:2328727
A;Accession: S09971
A;Molecule type: mRNA
A;Residues: 1-933 <KAS>
A;Cross-references: EMBL:X51730
R;Kastner, P.
submitted to the EMBL Data Library, February 1990
A;Reference number: S12464
A;Accession: S12464
A;Molecule type: mRNA
A;Residues: 1-343,'T',345-933 <KA2>
A;Cross-references: EMBL:X51730; NID:g35651; PIDN:CAA36018.1; PID:g35652
R;Misrahi, M.; Atger, M.; D'Auriol, L.; Loosfelt, H.; Meriel, C.; Fridlansky, F.; Gui
Biochem. Biophys. Res. Commun. 143, 740-748, 1987
A;Title: Complete amino acid sequence of the human progesterone receptor deduced from
A;Reference number: A03245; MUID:87184565; PMID:3551956
A;Accession: A03245
A;Molecule type: mRNA
A;Residues: 1-225,'G',227-255,'V',257-659,'V',661-933 <MIS
A;Cross-references: GB:M15716; NID:g189934; PIDN:AAA60081.1; PID:g189935
C;Genetics:
A;Gene: GDB:PGR
A;Cross-references: GDB:119493; OMIM:264080
A;Map position: 11q22.1-11q22.3
C;Superfamily: progesterone receptor; erbA transforming protein homology
C;Keywords: alternative splicing; DNA binding; nucleus; phosphoprotein; steroid hormo
F;1-933/Product: progesterone receptor form B #status predicted <MA1>
F;165-933/Product: progesterone receptor form A #status predicted <MA2>
F;565-829/Domain: erbA transforming protein homology <ERBA>
F;567-587/Region: zinc finger CCCC motif
F;603-627/Region: zinc finger CCCC motif
F;681-933/Domain: steroid binding #status predicted <STB>
F;41/Binding site: phosphate (Thr) (covalent) #status predicted
F;227,242,552,793/Binding site: phosphate (Ser) (covalent) #status predicted
F;329,374,601/Binding site: phosphate (Tyr) (covalent) #status predicted

  Query Match             16.1%;  Score 451;  DB 1;  Length 933;
  Best Local Similarity   27.2%;  Pred. No. 5.2e-27;
  Matches  131;  Conservative   81;  Mismatches  163;  Indels  106;  Gaps   18;

Qy          6 SPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHE------YPAMTF-------YSPA 52
              :|    :| :  |   :|| :     :||:   :          |||:          | |
Db        476 APPPCKAPGASGC---LLPRDG----LPSTSASAAAAGAAPALYPALGLNGLPQLGYQAA 528

Qy         53 VMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARS 112
              |:   :                | |  |   :| |                :|  :||
Db        529 VLKEGL----------------PQVYPPYLNYLRP----  -  -DSEASQSP       555

Qy        113 LEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAF 172
                  : |:| :|:                        | :| | ||| |||| :|  || |
Db        556 ---QYSFESLPQK1---  -  --  -   -    CLICGDEASGCHYGVLTCGSCKVF 590

Qy        173 FKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRQ 232
              |||:::| ::|:|   | | :|| |||:| |||||||  : |||  | :  ::   :  ||
Db        591 FKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRKCCQAGMVLGGRKFKK  FNKVRVV 648

Qy        233 RSADEQL--HCAGKAKRSGGHAPRVRELLLDALSPEQ-------LVLTLLEAEPPHVLIS 283
              |: |        |   | :|          || |        |:  |:  ||  :
Db        649 RALDAVALPQPLGVPNESQALSQRF------TFSPGQDIQLIPPLINLLMSIEPDVIYAG 702

Qy        284 RPSA-PFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMG 342
              :  | | :|:: || :| :::|: :: |:| :|||  | : ||: |::  || ::: |
Db        703 HDNTKPDTSSSLLTSLNQLGERQLLSVVKWSKSLPGFRNLHIDDQITLIQYSWMSLMVFG 762

Qy        343 LMWRSIDHPG--KLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCV 400
```

```
                 | |||  |     | |||||:|:    |        :   :       | :|::  :|:||:
Db          763 LGWRSYKHVSGQMLYFAPDLILNEQRMK-ESSFYSLCLTMWQIPQEFVKLQVSQEEFLCM 821

Qy          401 KAMILLNS-SMYPLVTATQ-DADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL 458
                | ::|||:  :  | : || .  | .  |. |     :  :  |  :|| |:  :| |
Db          822 KVLILINTIPLEGLRSQTQFEEMRSSYIRELIKA-----IGLRQKGVVSSSQRFYQLTKL 876

Qy          459 L 459
                |
Db          877 L 877


RESULT 32
QRMSN1
probable hormone receptor N10, nuclear - mouse
N;Alternate names: steroid hormone receptor homolog nur/77
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Jun-1999
C;Accession: S06953; A30059; A31844
R;Ryseck, R.P.; Macdonald-Bravo, H.; Mattei, M.G.; Ruppert, S.; Bravo, R.
EMBO J. 8, 3327-3335, 1989
A;Title: Structure, mapping and expression of a growth factor inducible gene encoding a
A;Reference number: S06953; MUID:90059925; PMID:2555161
A;Accession: S06953
A;Molecule type: DNA
A;Residues: 1-601 <RYS>
A;Cross-references: EMBL:X16995; NID:g53312; PIDN:CAA34862.1; PID:g53313
R;Hazel, T.G.; Nathans, D.; Lau, L.F.
Proc. Natl. Acad. Sci. U.S.A. 85, 8444-8448, 1988
A;Title: A gene inducible by serum growth factors encodes a member of the steroid and th
A;Reference number: A94211; MUID:89042200; PMID:3186734
A;Accession: A30059
A;Molecule type: mRNA
A;Residues: 1-601 <HAZ>
A;Cross-references: GB:J04113; NID:g200115; PIDN:AAA39843.1; PID:g200116
C;Genetics:
A;Map position: 15F
A;Introns: 295/3; 339/1; 389/3; 457/2; 517/1
C;Superfamily: probable hormone receptor N10, nuclear; erbA transforming protein homolog
C;Keywords: DNA binding, nucleus, steroid hormone receptor; transcription factor; zinc f
F;268-517/Domain: erbA transforming protein homology <ERBA>
F;270-290/Region: zinc finger CCCC motif
F;276 335/Domain: DNA binding #status predicted <DNA>
F;306-330/Region: zinc finger CCCC motif

  Query Match             16.0%;  Score 447.5;  DB 1;  Length 601;
  Best Local Similarity   28.9%;  Pred. No. 5.5e-27;
  Matches  158;  Conservative   63;  Mismatches  211;  Indels  115;  Gaps   20;

Qy            5 NSPSSLNSPSS----------YNCSQSIL--PLEHGSIYIPSSYVDSHHEYPAMTFYSPA 52
                :| ||  ||:|          | |    |  ||:       | |  |    |     ||:
Db           90 SSSSSATSPASASPKFEDPQVYGCYPGTLSGPLDETLSSSGSEYYGSPCSAP-----SPS 144

Qy           53 VMNYSIPSNVTNLEGGPGRQTTSPNV--LW------- --------PT------PGHLSP 88
                |:  || ::  :|  |  : |     ||                ||      |  ||
Db          145 TPNFQ-PSQLSPWDGSPGHFSPSQTYEGLWAWTEQLPKASSGPPPPPTFFSFSPPTGPSP 203

Qy           89 LVVHRQLSHLYAEPQKSPWCEARSLEHTLPV----------NRFT---IKRKVSGNRCAS 135
                :   |  |: |        | |  :::|           ||:|   |   |:  :  |
Db          204 SLAQSSLK-LPPPPATHQLGPGES--YSMPAAFPGLAPTSPNRDTSGILPAPVTSTKSRS 260

Qy          136 PVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDK 195
                  :|    |    |||| | ||  |||| :||||| ||||::|    ||| |   | :||
Db          261 GASGGSPDR----CAVCGDNASGLHYGVFTCEGCKGFFKRTVQKSAKYICLANKQCPVDK 316

Qy          196 NRRKSCQACRLRKCYEVGMVKCGSRREPCGYPLVPPQRSADEQLHCAGKAKRSGCHAPPV 255
                || || || :||  |||||           :||          |   | |
Db          317 RRRNRCQFCRFQKCLAVGMVK----------EVVR----------TDSLKGPRGPLPSK 355

Qy          256 RELLLDALSPEQIVITLIEAFPPHVISRPSAPFTFASMMMSLT---------------- 299
                 :   || ||  |: :|: |     | | ||    : |   |
Db          356 PKQPPDA-SPTNLITSLIRAH----IDSGPSTAKLDYSKFQFLVIPPFGKEDAGDVQQFY 410

Qy          300 KLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPD 359
                  |    |  : ||:||||||:||   ||  |||| ::|: :: | :||    |||||
Db          411 DLLSGSIDVIRKWAEKIPGFIELCPGDQDLILIESAFLELFILRLAYRSKPGEGKLIFCSG 470

Qy          360 LVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQD 419
                |||   : :|  |  :  | :|| :     | :    : |: |::          ||
Db          471 LVL--HQLQCARGFGDWIDNILAFSRSLHSLGVDVPAFACLSALV--------LITDRHG 520

Qy          420 ADSSKKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNM 479
                    |::  | | :   |   :|      |  |  |: ||  |  :|    :|:: :  :
Db          521 LQDPRRVEELQNRIASCLKEHMATVAGDPQPASC-LSRLLGKLPELRTLCTQGLQKIFCL 579

Qy          480 KCKNVVP 486
                | :::||
Db          580 KLEDLVP 586


RESULT 33
A29513
mineralocorticoid receptor - human
N;Alternate names: aldosterone receptor
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 20-Sep-1999
C;Accession: A29513
R;Arriza, J.L.; Weinberger, C.; Cerelli, G.; Glaser, T.M.; Handelin, B.L.; Housman, D
Science 237, 268-275, 1987
A;Title: Cloning of human mineralocorticoid receptor complementary DNA: structural an
A;Reference number: A29513; MUID:87263386; PMID:3037703
A;Accession: A29513
A;Molecule type: mRNA
A;Residues: 1-984 <ARR>
A;Cross-references: GB:M16801; NID:g187460; PIDN:AAA59571.1; PID:g407166
C;Genetics:
A;Gene: GDB:MLR
A;Cross-references: GDB:120188; OMIM:264350
A;Map position: 4q31-4q31
C;Superfamily: unassigned erbA-related proteins; erbA transforming protein homology
C;Keywords: DNA binding; transcription regulation; zinc finger
F;601-880/Domain: erbA transforming protein homology <ERBA>
F;603-624/Region: zinc finger
F;639-663/Region: zinc finger

  Query Match             16.0%;  Score 447.5;  DB 2;  Length 984;
  Best Local Similarity   25.5%;  Pred. No. 1e-26;
  Matches  150;  Conservative   89;  Mismatches  206;  Indels  143;  Gaps   21;

Qy            5 NSPSSLNSP-----SSYNCSQSI---------LPLEHGSIYIPSSYVDSHHEYPAMTFYS 50
                || ||  : |     : ::|| :           |   || :   |::|    |
Db          418 NSDSSPSVPIKQESTKHSCSGTSFKGNPTVNPFPFMDGSYF---SFMDDKDYYSLSGIIG 474

Qy           51 PAVMNY-------------------------SIPSN-VTNLEGGP- -----GRQTTSP 76
                | |  :                        ||||: :  :  |          | | |
Db          475 PPVPGFDGNCEGSGFPVGIKQEPDDGSYYPEASIPSSAIVGVNSGGQSFHYRIGAQGII- 533

Qy           77 NVLWPTPCHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNR-------ETLKREVS 129
                          ||      :  || : |  :    |:      |  |       | :  ||
Db          534 --------SLSRSARDQSPQHLSSPPPVNTLVESWKSHGDLSSRRSDGYPVLEYIPENVS 585

Qy          130 GNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPAIN 189
                  :   |  |  || | :  | || | ||| |||| :|  || |||||:::| ::|:|   |
Db          586 SSILRSVSI--GSSRPSKICLVCGDEASGCHYGVVTCGSCKVFFKRAVEGQHNYLCAGRN 643

Qy          190 QCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREPCGYRLVPRQRSADEQLHC-------- 241
                | ||| |||:| ||||:|| : ||   |:|: :   :| : :   :||
Db          644 DCIIDKIRRKNCPACRLQKCLQAGM-NLGARKSK---KLGKLKGTHEFQPQGQQPPPPPP 699

Qy          242 ------AGKAKRSGGHAPPVRELLLDALSPEQLVLI-----LLEAEPPHVLIS-- RPSAP 288
                     |   :    | |  :  ||    ||      :||   | :: :    | |
Db          700 PPQSPEEGTTYIAPAKEPSVNTALVPQLSIISRALIPSPVMVLENIEPEIVYAGYDSSKP 759

Qy          289 FTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSI 348
                |  ::: :| :|| |::: :: ||| :|||  | | ||: |::  || :    | |||
Db          760 DTAENLLSTLNRLAGKQMIQVVKWAKVLPGFKNLHIEDQITLIQYSWMCISSFALSWRSY 819
```

```
Qy        349 DHPGK--LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILL 406
              |    | |||||| : :|      : |:   :   : :|  |:|  :||  :| ::||
Db        820 KHINSQFLYFAPDLVFN-EEKMHQSAMYELCQGMHQISLQFVRLQLTFEEYTLMKVLLLL 878

Qy        407 NSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ--QSMR---LANLLML 461
              ::                          | | |: ||   : ||   :  | :
Db        879 ST-  - -- - -- -  - -- - -- - IPFDGLKSQAAFEEMRTNYIKELKKM 926

Qy        462 LSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML---- - ---NAHVLR 501
              ::   : | :   :   : | :  ::||: ::|          :| |:
Db        907 VTKCPNNSGQSWGRFYQL TKLLDSMHDLVSDLLEFCFYTFFESHALK 954


RESULT 84
150515
retinoid X receptor alpha - zebra fish
C;Species: Brachydanio rerio (zebra fish)
C;Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 02-Aug-2002
C;Accession: 150515
R;Jones, B.B.; Ohno, C.K.; Allenby, G.; Boffa, M.B.; Levin, A.A.; Grippo, J.F.; Petkovic
Mol. Cell. Biol. 15, 5226-5234, 1995
A;Title: New retinoid X receptor subtypes in zebra fish (Danio rerio) differentially mod
A;Reference number: A57301; MUID:96009547; PMID:7565671
A;Accession: 150515
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-441 <JON>
A;Cross-references: EMBL:U29940; NID:g1046294; PIDN:AAC59720.1; PID:g1046295
C;Genetics:
A;Gene: RXR
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: zinc finger
F;115-357/Domain: erbA transforming protein homology <ERBA>

  Query Match             15.3%;  Score 446;  DB 2;  Length 441;
  Best Local Similarity   27.4%;  Pred. No. 4.9e-27;
  Matches  139;  Conservative   79;  Mismatches  196;  Indels   90;  Gaps   16;

Qy         10 LNSPSSYNCSQSILPLEHGSIYIP------SSYVDSHHEYPAMTFYSPAVMN---- YSI 58
              :::  :|    | | : || :  |       | |  ||          |:::|     ||:
Db          1 MDNNDTYLHLSSSLQVAHGHLSSPPSQPPLGSMVSHHH- --   ---PSLINGLGSPYSV 51

Qy         59 PSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPV 118
              :|:   |      |||:  ||  ::      ::          || |    |    |
Db         52  - ITSSSLG-----SPSASMPTTSNMGYGALN-------SPQMNSLNSVSSSEDIKP 94

Qy        119 VNRETLKRKVSGNGCASPVEGPGSFPDAHPCAVCSDYASGYHYGVWSCEGCKAFFKRSIQ 178
                  :    | ||  |  ||   | ||:| | :|| ||||:|||||| ||||:|:
Db         95  - PPGLAGLGSYPCGSP GSLSK--HICAICGDRSSGKHYGVYSCEGCKGFFKRTIR 146

Qy        179 GHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREPCGYRLVRPQPSADEQ 238
              | |     | ||| :|  || || :||  :|| :   : ||      | :  :| :
Db        147 KDLTYTCRDNKDCQIDKRQRNRCQYCRYQKCLAMGMKREAVQEERQ----QRGKEPSGNE 201

Qy        239 LHCAGKAKRSGGHAPKVKELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMM---- 294
              : :                    : | |  | :|:||   :  :   :  : |:||
Db        202 VDSS                  SSFNEEMPVEKILDAE-- LAVEPKTLAYMLSSMSNGI 240

Qy        295 --MMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHP 351
              : :: : |||:|  :: |||:|| | :| | ||| || : | |:|:     ||:
Db        241 NDPVTNICQAADKQLFTLVEWAKRIPHFSDLPLDDQVILLRAGWNELLIASFSHRSVTVK 300

Qy        352 GKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSSM 410
              ::| | : |        |:  ||| :|    |: |::::   |  |::|::| |
Db        301 DGILLATGLHVHRSSAHSA-GVGSIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNPDA 359

Qy        411 YPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASN 470
              |   ::      |:  |  |              :  | | ||: |  :|
Db        360 KGLSNPSEVEALREKVYASLEAYCKH---- --  --NYPDQFGRFAKLLLRLPALRSIGL 409

Qy        471 KGMEHLLNMKCKNVVPVYDLLLEMLNA 497
              | :|||   |     |:   |:||| |
Db        410 KCLEHLFFFKLIGDTPIDTFLMEMLEA 436


RESULT 85
A47278
retinoid X receptor alpha - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
C;Accession: A47278
R;Gearing, K.L.; Gottlicher, M.; Teboul, M.; Widmark, E.; Gustafsson, J.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1440-1444, 1993
A;Title: Interaction of the peroxisome-proliferator-activated receptor and retinoid X
A;Reference number: A47278; MUID:93165712; PMID:8381967
A;Accession: A47278
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-467 <CEA>
A;Cross-references: GB:L06482; NID:g206818; PIDN:AAA42093.1; PID:g206819
A;Note: sequence extracted from NCBI backbone (NCBIN:124909, NCBIP:124911)
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: DNA binding; transcription regulation; zinc finger
F;138-383/Domain: erbA transforming protein homology <ERBA>

  Query Match             15.8%;  Score 442.5;  DB 2;  Length 467;
  Best Local Similarity   28.5%;  Pred. No. 9.8e-27;
  Matches  143;  Conservative   70;  Mismatches  208;  Indels   81;  Gaps   15;

Qy         87 VDSHHEYPAMTFYSPAVMNYSIPSNVINLEGGPGRQTTSPNVLWPT- PGHLSPL- - V 90
              :|: |  |          :::|   | ::|    || : :   | |:  ||  |||
Db          1 MDFKHELP---------LDFSTQVNSSSISSPTGRGSMAAPSLHPSLGPGLGSPLGSPGQ 51

Qy         91 VHRQLSHLYA--EPQKSPWCEARSL--EHTLPV NKETLKRKVSGNRCASPVIGPGSKRD 145
              :|  :| | :          |:    |    ||: |    || :    :  ||:    | |
Db         52 LHSPTSTLSSPTNGMGPPFSVISSPMGPHSMSVPTTPTLGFETGSPQLNSPMNPVSSSED 111

Qy        146 -------------------------AHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
                                        | ||:| | :|| ||||:|||||| ||||:::
Db        112 IKPPLGLNGVLKVPAHPSGNMSSFTKHICAICGDRSSGKHYGVYSCEGCKGFFKRTIVRKD 171

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREPCGYRLVRPQRSADEQLE 240
                | |     | ||| :|  || || :||  :|| :   : |        ||| |   :
Db        172 LTYTCRDNKDCLIDKRQRNRCQYCRYQKCLAMGMKREAVQEE------- -PQRCKPRNEN 223

Qy        241 CAGKAKRSGGHAPEVPELLLDALSPEQLVLTLLEA----EPPHVLISRPSAPFTEASMMM 296
                          :   | | ::|   |: |     | :||     |     | |: | |
Db        224 EVESTSSANEDMP VEKILEAELAVEPKTETYVEANMGLNP-----SSPNDPVT------ 271

Qy        297 SLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIF 356
              :: : |||:|  :: |||:|| | :| | ||| || : | |:|:     |||      ::
Db        272 NICQAADKQLFTLVEWAKRIPHFSELPLDDQVILLRAGWNELLIASFSHRSIAVKDGILL 331

Qy        357 APDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSSMYPLVT 415
              |  | : |:       |:  ||| :|    |: |::::   |  |::|::| |     |
Db        332 ATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNPDSKGLSN 390

Qy        416 ATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEH 475
              :      |:   | |                    :|  | | ||: |  :|    | :||
Db        391 PAEVEALREKVYASLEAYCKH---- --  --KYPEQFGRFAKLLLRLPALRSIGLKCLEH 440

Qy        476 LLNMKCKNVVPVYDLLLEMLNA 497
              |   |      |:   |:||| |
Db        441 LFFFKLIGDTPIDTFLMEMLEA 462


RESULT 86
S26668
retinoic acid receptor X-alpha - mouse
N;Alternate names: 9-cis retinoic acid receptor; retinoid X receptor alpha
C;Species: Mus musculus (house mouse)
C;Date: 14-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Aug-2002
C;Accession: S26668; C41727; S54450
R;Mangelsdorf, D.J.; Borgmeyer, U.; Heyman, R.A.; Zhou, J.Y.; Ong, E.S.; Oro, A.E.; K
Genes Dev. 6, 329-344, 1992
```

```
A;Title: Characterization of three RXR genes that mediate the action of 9-cis retinoic a
A;Reference number: S26668; MUID:92192447; PMID:1312497
A;Accession: S26668
A;Molecule type: mRNA
A;Residues: 1-467 <MAN>
A;Cross-references: EMBL:X66223; NID:g54021; PIDN:CAA46962.1; PID:g54022
A;Experimental source: tissue type embryo
R;Leid, M.; Kastner, P.; Lyons, R.; Nakshatri, H.; Saunders, M.; Zacharewski, T.; Chen,
Cell 68, 377-395, 1992
A;Title: Purification, cloning, and RXR identity of the HeLa cell factor with which RAR
A;Reference number: A41727; MUID:92127595; PMID:1310259
A;Accession: C41727
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-466 <LEI>
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:77548)
R;Tontonoz, P.; Graves, R.A.; Budavari, A.I.; Erdjument-Bromage, H.; Lui, M.; Hu, E.; Te
Nucleic Acids Res. 22, 5628-5634, 1994
A;Title: Adipocyte-specific transcription factor ARF6 is a heterodimeric complex of two
A;Reference number: S53748; MUID:95140625; PMID:7838715
A;Accession: S54450
A;Molecule type: protein
A;Residues: 5-16;27-46;291-307;446-465 <TON>
A;Experimental source: adipose tissue
C;Comment: Composed of an N-terminal domain (needed to activate transcription), a conser
C;Complex: homodimer; heterodimers with 9-cis retinoic acid (RXR), thyroid (TRP), vitami
C;Function:
A;Description: transcription regulation
A;Pathway: retinoic acid response pathway
A;Note: homodimers recognize palindromic sequences, responds to a variety of endogenous
t subunit, and apparently do not require 9-cis retinoic acid receptor to bind the ligand
ocl in the absence of ligand and repress transcription (interaction with a corepressor p
C;Superfamily: retinoic acid receptor alpha; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; receptor; transcription regulation; zinc finger
F;138-483/Domain: erbA transforming protein homology <ERBA>
F;140-160/Region: zinc finger CCCC motif
F;176-200/Region: zinc finger CCCC motif

  Query Match             15.8%;  Score 442.5;  DB 2;  Length 467;
  Best Local Similarity   28.3%;  Pred. No. 9.8e-27;
  Matches  150;  Conservative   70;  Mismatches  209;  Indels  101;  Gaps   18;

Qy          1 MDIKN----------SPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVD------------ 38
              || |:          : ||||||:             ||: :|| :
Db          1 MDTKHFLPLDFSTQVNSSSLNSPTG----------RGSMAVPSLHPSLGPGIGSPLGSP 49

Qy         39 SHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQT--TSPNVLWPI--PGHLSPLVVHRQ 94
                 |  | || :     | :| :   ||   :  |:| : : | |   ||:
Db         50 GQLHSPISTLSSP-INGMGPPFSVISSPMGPHSMSVPTTPTLGFGTGSPQLNSPM----- 104

Qy         95 LSHLYAFPQKSPWCEARSLEHTLPVNRETLK--RKVSGNRCASPVTGPGSKRDAHFCAVC 152
                      :|      ..  | .|   ||    |||  .                | ||.|
Db        104 ----------NPVSSTEDIKPPLGLN-GVLKVPAHPSGNMASF---------TKHICAIC 143

Qy        153 SDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEV 212
                | :|| ||||:|||||| ||||:::    | |      | ||| :|  || || :|| :
Db        144 GDRSSGKHYGVYSCEGCKGFFKRTVRKDLTYTCRDNKDCLIDKRQRNRCQYCRYQKCLAM 203

Qy        213 GMVKCGSRRERCGYRLVPRQRSADFQLHCACKAKRSCGHAPFVRELLLDALSPEQLVLTL 272
                || :   : |        ||| |   :       :   | | :::|   |: |    |
Db        204 GMKREAVQEE--------RQRGKDPNENEVESTSSANEDMP-VEKILEAELAVEPKTETY 254

Qy        273 LEA----EPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQV 328
                :||      |      | |: | |       :: : |||:|  :: |||:|| | || | |||
Db        255 VEANMGLNP-----SSPNDPVT------NICQAADKQLFTLVEWAKRIPHFSELPLDDQV 303

Qy        329 RLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRF 387
                 || : | |:|·   |||      ·· | · · |·       |·  ||| ·| |·
Db        304 ILLRAGWNELLIASFSHRSIAVKDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKM 362

Qy        388 RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGIS 447
                |::::   |  |::|:::| |     |    :      |:   | |

Db        363 RDMQMDKTELGCLRAIVLFNPDSKGLSNPAEVEALREKVYASLEAYCKH          K 412

Qy        448 SQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
                 :|  | | ||: |  :|    | :|||   |     |:   |:|| |
Db        413 YPEQPGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 462


RESULT 37
A37251
probable nuclear hormone receptor NAK1 - human
N,Alternate names: early response protein NAK1, nuclear hormone receptor SI-59; stero
C;Species: Homo sapiens (man)
C;Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 20-Jan-2000
C;Accession: A37251; A61191; A61451
R;Nakai, A., Kartha, S., Sakurai, A., Toback, F.G., DeGroot, L.J.
Mol. Endocrinol. 4, 1438-1443, 1990
A;Title: A human early response gene homologous to murine nur77 and rat NGFI-B, and r
A;Reference number: A37251; MUID:91133413; PMID:2283997
A,Accession: A37251
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-598 <NAK>
A;Cross-references: GB:D49728; NID:g1813881; PIDN:BAA08565.1; PID:g1813882
A,Note: it is uncertain whether Met-1 or Met-35 is the initiator
R;Bondy, G.P.
Cell Growth Differ. 2, 203-208, 1991
A;Title: Phorbol ester, forskolin, and serum induction of a human colon nuclear hormo
A;Reference number: A61191; MUID:91329278; PMID:1651101
A;Accession: A61191
A;Molecule type: mRNA
A;Residues: 25-252,'I',254-598 <BON>
R;Chang, C.; Kokontis, J.; Liao, S.; Chang, Y.
J. Steroid Biochem. 34, 391-395, 1989
A;Title: Isolation and characterization of human TR3 receptor: a member of steroid re
A;Reference number: A61451; MUID:90173205; PMID:2626032
A,Accession: A61451
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-261,'P',263-469,'L',471-598 <CHA>
A;Cross-references: GB:L13740; NID:g292833; PIDN:AAA36763.1; PID:g292834
C;Genetics:
A;Gene: GDB:HMR
A;Cross-references: GDB:125357; OMIM:139139
A;Map position: 12q13-12q13
C;Superfamily: probable hormone receptor N10, nuclear; erbA transforming protein homo
C;Keywords: DNA binding; hormone receptor; nucleus; transcription regulation; zinc fi
F;265-514/Domain: erbA transforming protein homology <ERBA>

  Query Match             15.8%;  Score 442.5;  DB 2;  Length 598;
  Best Local Similarity   28.8%;  Pred. No. 1.3e-26;
  Matches  152;  Conservative   58;  Mismatches  202;  Indels  115;  Gaps   17;

Qy          2 DIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTF------------ 48
              |   || |  |||: :    |    ||           |  |: |:
Db        130 DYYGSPCSAPSPSTPSFQPPQLSPWDGSF---------GHFSPSQIYEGLKAWTEQLPKA 180

Qy         49 ----YSPAVMNYSIPSNVTNLEGGPGRQ-TTSPNVLWPTPGHLSPLVVHRQLSHLYAFPQ 103
                   ||  ::| |:       ||      ||  |:|:           | :|   |
Db        181 SGPPQPPAFFSFSPPT-------GPSPSLAQSPLKLFPS----------QAIHQLGEG- 221

Qy        104 KSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVT-----GPGSKRDAHFCAVCSDYA 156
                  |: |:    |   |          :|||       ||      |||| | |
Db        222 -----ESYSMPTAFPGLAPTSPHLEGSGILDTPVTSTKARSGAPGGSEGR--CAVCGDNA 274

Qy        157 SGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVK 216
                | |||| :||||| ||||::| :  ||| |   |  :|| ||  || || :||  |||||
Db        275 SCQHYGVRTCEGCKGFFKRTVQKNAKYICLANKDCPVDKRRRNRCQFCRFQKCLAVGMVK 334

Qy        217 CGSPRERCGYRLVPEQPEADQLHCACKAKRSCGHAPFVRELLLDALSPEQLVLTLLEAE 276
                          :||                |  | |  :  || ||  |: :|: |
Db        335 ----------EVVR----------TDSLKGRRGRLPSKPKQPPDA-SPANLLTSLVRAH 372

Qy        277 PPHVLISRPSAPFTEASMMMSLT----  ----------KLADKELVHMISWAKKIPGFV 320
```

```
                    | | ||     : |    |                        |    |  :  ||:||||
Db          473    --LDSGPSTAKLDYSKFQELVLPHFGKEDAGDVQQFYDLLSGSLEVIRKWAEKIPGFA 428

Qy          321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDML 380
                ||| ||  |||| :|: :: | :||    ||:||   ||| | :  |  |  :  | :
Db          429 ELSPADQDLLLESAFLELFILGLAYRSKPGEGKLIFCSGLVLHRLQ--CARGFGDWIDSI 486

Qy          381 LATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWV 439
                || :     | :    : |: |::        |:|        |::  | | :   |
Db          487 LAFSRSLHSILVDVPAFACISALV- ------LITDRHGLQEPPRVEELQNRIASCLEEH 538

Qy          440 VIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVP 486
                | | :|    | :  |: || |  :|    ::|: :  :| :::||
Db          539 VAAVAG--EPQPASCLSRLLGKLPELRTLCTQGLQRIFYLKLEDLVP 583


RESULT 38
A35466
progesterone receptor form B - chicken
N;Contains: progesterone receptor form A
C;Species: Gallus gallus (chicken)
C;Date: 18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C;Accession: A35466; S06284; A40903; A24661; A24312; A40911; A61552
R;Jeltsch, J.M.; Turcotte, B.; Garnier, J.M.; Lerouge, T.; Krozowski, Z.; Gronemeyer, H.
J. Biol. Chem. 265, 4967-4974, 1990
A;Title: Characterization of multiple mRNAs originating from the chicken progesterone re
A;Reference number: A35466; MUID:90154085; PMID:2403488
A;Accession: A35466
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-786 <JEL>
A;Cross-references: GB:M32742; GB:J05240; NID:g212554; PIDN:AAA49011.1; PID:g212558
R;Gronemeyer, H.; Turcotte, B.; Quirin-Stricker, C.; Bocquel, M.T.; Meyer, M.E.; Krozows
EMBO J. 6, 3985-3994, 1987
A;Title: The chicken progesterone receptor: sequence, expression and functional analysis
A;Reference number: S06284; MUID:88166640; PMID:3443098
A;Accession: S06284
A;Molecule type: DNA
A;Residues: 1-786 <GRO>
A;Cross-references: EMBL:Y00092; NID:g63744; PIDN:CAA68282.1; PID:g63745
R;Conneely, O.M.; Dobson, A.D.W.; Tsai, M.J.; Beattie, W.G.; Tott, D.O.; Buckaby, C.S.;
Mol. Endocrinol. 1, 517-525, 1987
A;Title: Sequence and expression of a functional chicken progesterone receptor.
A;Reference number: A40903; MUID:91042592; PMID:3153474
A;Accession: A40903
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-64,'E',65-786 <CON>
A;Cross-references: GB:M37518
R;Conneely, O.M.; Sullivan, W.P.; Tott, D.O.; Birnbaumer, M.; Cook, R.G.; Maxwell, B.L.;
Science 238, 767-770, 1986
A;Title: Molecular cloning of the chicken progesterone receptor.
A;Reference number: A24661; MUID:86289413; PMID:2426779
A;Accession: A24661
A;Molecule type: mRNA
A;Residues: 128-133,'E',135-147,'E',149-164 <CO2>
A;Note: the authors translated the codon CAG for residue 7 as Glu and CAG for residue 21
R;Jeltsch, J.M.; Krozowski, Z.; Quirin-Stricker, C.; Gronemeyer, H.; Simpson, R.J.; Garn
Proc. Natl. Acad. Sci. U.S.A. 83, 5424-5428, 1986
A;Title: Cloning of the chicken progesterone receptor.
A;Reference number: A24312; MUID:86287271; PMID:2426697
A;Accession: A24312
A;Molecule type: mRNA
A;Residues: 417-490 <JE2>
A;Cross-references: GB:M14280; NID:g212607; PIDN:AAA49039.1; PID:g212608
A;Note: amino acid and corresponding nucleotide sequences are also shown for three small
R;Birnbaumer, M.; Hinrichs-Roscello, M.V.; Cook, R.G.; Schrader, W.T.; O'Malley, B.W.
Mol. Endocrinol. 1, 249-259, 1987
A;Title: Chemical and antigenic properties of pure 108,000 molecular weight chick proges
A;Reference number: A40911; MUID:88288199; PMID:3453892
A;Accession: A40911
A;Status: preliminary
A;Molecule type: protein
A;Residues: 128-133,'E',135-147,'E',149-164;546-558 <BIR>
R;Simpson, R.J.; Grego, B.; Govindan, M.V.; Gronemeyer, H.
Mol. Cell. Endocrinol. 52, 177-184, 1987
A;Title: Peptide sequencing of the chick oviduct progesterone receptor form B.
A;Reference number: A61552; MUID:88005426; PMID:3653503
A;Accession: A61552
A;Molecule type: protein
A;Residues: 136-153;168-174;195-228;526-547,'X',549;546-563 <SIM>
C;Genetics:
A;Introns: 490/2; 451/1; 490/1; 591/1; 639/2; 683/1; 735/3
C;Superfamily: progesterone receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; zinc finger
F;1-786/Product: progesterone receptor form B #status predicted <MA1>
F;128-786/Product: progesterone receptor form A #status predicted <MA2>
F;419-682/Domain: erbA transforming protein homology <ERBA>
F;421-441/Region: zinc finger
F;457-481/Region: zinc finger

  Query Match             15.8%;  Score 442.5;  DB 2;  Length 786;
  Best Local Similarity   25.3%;  Pred. No. 1.9e-26;
  Matches  129;  Conservative   92;  Mismatches  178;  Indels  111,  Gaps   17;

Qy            7 PSSLNSPSSYNCSQSILPLEHGSIYIPSS----  -YVQSHHEYPAMTPYSPA---VMN 55
                |     | :|   || | ::   : :|::          :    |      |:   |:
Db          275 PDGKEGPFAYGEFQSALKIKEEGVGLPAAPPPFIGAKAAPADFAQPPRAGQEPSLECVLY 334

Qy           56 YSIPSNVINLEGGPGRQTTSPNVLWPTPGHLSPL- --            VVHRQLSHL-   98
                : |  :   | |    : |:     || |||                 |:  |  |
Db          335 KAEPPLLPGAYGPPAAPDSLPSTSAAPPGLYSPIGLNGHHQALGFPAAVLKEGLPQLCPP 394

Qy           99 ---YAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDY 155
                   |  |       ::| ::::     |:| :|:                    | :| |
Db          395 YLGYVRPDTE---TSQSSQYSF----ESLPQKI--  -    ----- --- ---CLICGDE 427

Qy          156 ASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMV 215
                ||| |||| :|  || ||||:::| ::|:|   | | :|| |||:| |||||||  : ||
Db          428 ASGCHYGVLTCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRKCCQAGMV 487

Qy          216 KCGSRRERGG-YRLVPRQPSADFQLHCAGKAKRSGGHAPRVRELLLDAL  --- SPEQL 268
                  | : ::    ::||    | :|              | | :    :|      || |
Db          488 LGGRKFKKLNKMKVVRTLDVALQQ-------------PAVLQDETQSLFQRLSFSPNQE 533

Qy          269 V------LTLLEAEPPHVLIS--RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFV 320
                :      :::|    | |: :    : | | :|:: ||  | :::|: :: |:| :|||
Db          534 IPFVPPMISVLRGIEPEVVYAGYDNTKPETPSSLLTSLNHLCERQLLCVVKWSKLLPGFR 593

Qy          321 ELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGK----  -CVE 371
                 | : ||: :::  || :::  : |||  |     | ||||||:|:    |       |:
Db          594 NLHIDDQITLIQYSWMSLMVFAMGWRSYKHVSGQMLYFAPDLILNEQRMKESSFYSICLS 653

Qy          372 GILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRK--LAHL 429
                    :     |  |::  :|:||:||::|||:     :  :    |  |   :  |
Db          654 -    --MWQLPQEFVRLQVSQEEFLCMKALLLLNTIPLEGLRSQSQFDEMRTSYIREL 705

Qy          430 LNAVTDALVWVIAKSGISSQQQSMRLANLL 459
                : |     : :  |  ::: |:  :|  |:
Db          706 VKA ----IGLRQKGVVANSQRFYQLTKLM 740


RESULT 39
S44047
glucocorticoid receptor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Jan-1995 #sequence_revision 27-Oct-1995 #text_change 22-Jun-1999
C;Accession: S45348; S44047; S42083
R;Gao, X.; Kalkhoven, E.; Peterson-Maduro, J.; van der Burg, B.; Destree, O.H.J.
Biochim. Biophys. Acta 1218, 194-198, 1994
A;Title: Expression of the glucocorticoid receptor gene is regulated during early emb
A;Reference number: S45348; MUID:94289478; PMID:8018720
A;Accession: S45348
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-776 <GAO>
```

```
A;Cross-references: EMBL:X72211
A;Experimental source: embryonic stage 17
R;Gao, X.; Kalkhoven, E.; Peterson-Maduro, J.; van der Burg, B.; Destree, O.H.J.
submitted to the EMBL Data Library, May 1993
A;Description: The expression of glucocorticoid receptor gene is regulated durin; early
A;Reference number: S44047
A;Accession: S44047
A;Molecule type: mRNA
A;Residues: 1-271,'P',273-776 <GAW>
A;Cross-references: EMBL:X72211; NID:g444042; PIDN:CAA51010.1; PID:g444043
A;Experimental source: embryonic stage 17
R;Picard, D.
submitted to the EMBL Data Library, February 1994
A;Reference number: S42083
A;Accession: S42083
A;Molecule type: mRNA
A;Residues: 452-501,'P',503,'A',505,'A',507-510,'N',512-518,'M',520-543,'F',545-550,'M',
A;Cross-references: EMBL:X77764; NID:g456373; PIDN:CAA54804.1; PID:g456374
A;Experimental source: liver
C;Superfamily: glucocorticoid receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation; <E
F;418-673/Domain: erbA transforming protein homology <ERBA>
F;420-440/Region: zinc finger CCCC motif
F;456-480/Region: zinc finger CCCC motif

  Query Match             15.7%;  Score 441.5;  DB 1;  Length 776;
  Best Local Similarity   39.2%;  Pred. No. 2.3e-26;
  Matches  126;  Conservative   61;  Mismatches  153;  Indels   77;  Gaps   14;

Qy        118 PVNRETLKRKVSGNRCASP--------------VTGPGSKRDAHPCAVCSDYASGYHYG 162
              | |     | |  |  :||                 |||  |     | |||| ||| |||
Db        378 PGNVNFPNRSVFSNGYSSPGIRSDASPSPSTSSTSTGPPPK----LCLVCSDEASGCHYG 433

Qy        163 VWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLPKCYEVGMVKGGIRKE 222
              | :|  || |||::::| ::|:|   | | ||| |||:| ||| ||| : ||    :|:
Db        434 VLTCGSCKVFFKRAVEGQHNYLCAGRNDCIIDKIRRKNCPACRYRKCLQAGM-NLEARKT 492

Qy        223 RGGYRIVPRQPSADFQLHCAGKAKRSGGHAPPVRELL---LDALSPEQLVLTLLEAEPPH 279
              :   : :::  :|          |       | |:   :  |:|   :::|||   |
Db        493 KKKIKGIQQSTTA---------TARESPETSMTRTLVPASVAQLTP--TLISLLEVIEPE 541

Qy        280 VLIS--RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWME 337
              || |    | | |   :| ||  |  :::|  : ||| ||||  | | ||: ||:  ||
Db        542 VLYSGYDSSIPDTTRRLMSSLNMLGGRQVVSAVKWAKAIPGFRNLHLDDQMTLLQYSWME 601

Qy        338 VLMMGLMWRSIDHPGK--LIFAPDLVLDRD-------EGKCVEGILEIFDMLLATTSRFR 388
              :::  | |||        | ||||||:  |       : :| |       :|
Db        602 IMVFALGWRSYKQTNGSILYFAPDLVITEDRMHLPFMQEVCQE--------MLKIAGEMS 654

Qy        389 ELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAH--LLNAVTDALVWVIAKSGI 446
              |:: : ||||:| ::|       : |  ::   |  |     :  : :    :: : |
Db        654 SLQISYDEYLCMKVLLL-------MCTIPKEGLKSHALFEEIRMTYIKELGKAIVKREGN 709

Qy        447 SSQ--QQSMRLANLI-------------MLISHVRHASNKGMEHLLNMKCKNVVPVY 488
              |||  |:  :|  ||               || :  : :      :|:    | :| |
Db        707 SSQNWQRFYQLTKLLDSMHEVAENLLAFCFLSFLDKSMSIEFPEMLSEIISNQIPKY 763


RESULT 40
A41401
mineralocorticoid receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 20-Sep-1999
C;Accession: A41401
R;Patel, P.D.; Sherman, T.G.; Goldman, D.J.; Watson, S.J.
Mol. Endocrinol. 3, 1877-1885, 1989
A;Title: Molecular cloning of a mineralocorticoid (type I) receptor complementary DNA fr
A;Reference number: A41401; MUID:90114194; PMID:2558305
A;Accession: A41401
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-981 <PAT>
A;Cross-references: GB:M36074; NID:g205340; PIDN:AAA41583.1; PID:g205341
C;Superfamily: unassigned erbA related proteins; erbA transforming protein homology
C;Keywords: DNA binding; transcription regulation; zinc finger
F;602-877/Domain: erbA transforming protein homology <ERBA>
F;604-624/Region: zinc finger
F;644-664/Region: zinc finger

  Query Match             15.7%;  Score 441.5;  DB 2;  Length 981;
  Best Local Similarity   26.0%;  Pred. No. 3e-26;
  Matches  149;  Conservative   87;  Mismatches  205;  Indels  131;  Gaps   21;

Qy          5 NSPSSLNSP---------SSYNCSQST---------LPLFHGSIYIPSSYVDSHHEYPAM 46
              ||  | :||          | ::|| :          |   || :   |::|    |
Db        415 NSKISPSSPFSVPIKQESSKHSCSGASFKGNPTVNPFPPMDGSYF---SFMDDKDYYSLS 471

Qy         47 TFYSPAVMNY-------------------------SIPSN-VINLEGGP-------GRQ 72
                | |  :                          ||||: :  :  |        | |
Db        472 GILGPFVPGFDGSCEDSAFPVGIKQEPMDGSYYFEASIPSSAIVGVNSGGQSFHYRIGAQ 531

Qy         73 TTSPNVLWPTPGHLSPLVVHKQLSHLYAEPQKSPWCEARSLEHTLPVNR----- - ETLK 125
               |              ||      :  || : |  :   |:       |  |       | :
Db        532 GII          SLSSSPFPQSPQHLSSFPPVNTLVESWEPHCPLSSFFSDCYFVLEYIP 582

Qy        126 PEVSNPCASPVTGPGSKEDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYIC 185
              || :   |  |  || | :  | || | ||| |||| :|  || ||||:::| ::|:|
Db        583 ENVSSSTLRSVSI--GSSKFSKICLVCGDEASGCHYGVVTCGSCKVFFKRAVEGQHNYLC 640

Qy        186 PATNQCTIDKNRRKSCQACRLKKCYEVGMVKGGIRKDRRGGRMLKHKREQDGEGRGEVGS 245
              | | ||| |||.| ||||.|| : ||   |:|: :     |    : |       :
Db        641 AGRNDCIIDKIRRKNCPACRLQKCLQAGM-NLGARKSKKLGKIKGLHEEQPQQPPPPPQ 699

Qy        234 SADEQLHCAGKAKRSGGHAPRVREL--LLDALSPEQLVLTLLEAEPPHVLISRP-SAPFT 290
              | :|        |    ::  | :|  :  ||:|   :: |  ||  |       | | |
Db        700 SPEEGITYIAPTKEPSVNSALVPQLTSITHALTPSPAMI-LENIEPETVYAGYDNSKPDT 758

Qy        291 EASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDH 350
              |:: :| :|| |::: :: ||| :|||  | | ||: |:   ||  :    | |||  |
Db        759 AESILLSTLNRLAAKQMIQVVKWAKVLPGFKNLPLEDQITLIQYSWMCLSSFALSWRSYKH 818

Qy        351 PGK--LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKIQHKEYLCVKAMILLNS 408
              | |||||  : :|     : |:   :   : :|  |:|  :||  :| ::|::::
Db        819 INSQLLYFAPDLVFN-EEKMHQSAMYELCQGMRQISLQFVKLQLTFEEYSIMKVLLLLST 877

Qy        409 SMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ--QSMR---LANLLMLLS 463
                                          : | : ||   : ||   :  | :::
Db        878 -------------------------------VPKDGLKSQAAFEEMRTNYIKELRKMVT 905

Qy        464 HVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                :|| :  :    :  | :  ::|||: ::|
Db        906 KCPNSSGQSWQRFYQLTKLLDSMHDLVSDLL 936


Search completed: June 16, 2003, 13:16:39
Job time : 30.4238 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 07:39:07 ; Search time 13.4083 Seconds
(without alignments)
1639.467 Million cell updates/sec

Title: US-08-826-361A-25
Perfect score: 2805
Sequence: 1 MDIKNSPSSLNSPSSYNCSQ. ECSPAEDSKSKECSQNPQSQ 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2805	100.0	530	1	ESR2_HUMAN	Q92731 homo sapien
2	2615	93.2	530	1	ESR2_CALJA	Q95171 callithrix
3	2496	89.0	530	1	ESR2_RAT	Q62986 rattus norv
4	2495	88.9	530	1	ESR2_MOUSE	O08537 mus musculu
5	2463.5	87.8	527	1	ESR2_BOVIN	Q9xsb5 bos taurus
6	2456.5	87.5	527	1	ESR2_SHEEP	Q9tu15 ovis aries
7	2415	86.1	526	1	ESR2_PIG	Q9xsw2 sus scrofa
8	2101	74.9	554	1	ESR2_STRVU	Q9ppp2 strongylo
9	1959	69.8	472	1	ESR2_COTJA	O93511 coturnix co
10	1948.5	69.5	472	1	ESR2_CHICK	Q9ptu5 gallus gall
11	1549	55.2	573	1	ESR2_ANGJA	O13012 anguilla ja
12	1466	52.3	568	1	ERR1_CARAU	Q9w669 carassius a
13	1440.5	51.4	568	1	ESR2_ONCMY	P57782 oncorhynchu
14	1441	51.0	673	1	ESR2_MICUN	P57781 micropogoni
15	1398	49.8	279	1	ESR2_MACMU	Q9tte5 macaca mula
16	1385.5	49.4	559	1	ESR2_SPAAU	Q9w6m2 sparus aura
17	1381	49.2	575	1	ESR2_ICTPU	Q9iak1 ictalurus p
18	1378.5	49.1	610	1	ERR2_CARAU	Q9ial9 carassius a
19	1373	48.9	565	1	ESR3_MICUN	P57783 micropogoni
20	1352	48.2	557	1	ESR2_OPENI	Q9yh32 oreochromis
21	1233.5	44.0	595	1	ESR1_HUMAN	P03372 homo sapien
22	1228.5	43.8	587	1	ESR1_DROME	Q01250 drosophila
23	1225.5	43.7	589	1	ESR1_CHICK	P06212 gallus gall
24	1220	43.5	594	1	ESR1_HORSE	Q9tv98 equus caball
25	1219.5	43.5	595	1	ESR1_PIG	Q29040 sus scrofa
26	1219	43.5	595	1	ESR1_MESAU	Q9qzj5 mesocricetu
27	1218.5	43.4	599	1	ESR1_MOUSE	P19785 mus musculu
28	1216.5	43.4	600	1	ESR1_RAT	P06211 rattus norv
29	1214	43.3	585	1	ESR1_OPENI	Q9yh33 oreochromis
30	1214	43.3	586	1	ESR1_XENLA	P81559 xenopus lae
31	1204	42.9	569	1	ESR1_BRARE	P57717 brachydanio
32	1204	42.9	579	1	ESR1_SPAAU	Q9pvc9 sparus aura
33	1199	42.7	620	1	ESR1_ORYLA	P50241 oryzias lat
34	1194.5	42.6	581	1	ESR1_PAGMA	O42132 pagrus majo
35	1184.5	42.2	583	1	ESR1_OREAU	P50240 oreochromis
36	1177	42.0	525	1	ESR1_MICUN	P57753 micropogoni
37	1165	41.5	622	1	ESR1_ONCMY	P16058 oncorhynchu
38	1160.5	41.4	617	1	ESR1_ICTPU	Q9yhz7 ictalurus p
39	1134.5	40.4	535	1	ESR1_SALSA	P50242 salmo salar
40	877	31.3	307	1	ESR1_CNEUN	Q91424 cnemidophor
41	799.5	28.5	349	1	ESR1_ANOCA	Q9yht3 anolis caro
42	648	23.1	458	1	ERR3_HUMAN	O75454 homo sapien
43	631.5	22.5	433	1	ERR2_RAT	P11475 rattus norv
44	630	22.5	500	1	ERR2_HUMAN	O95718 homo sapien
45	619.5	22.1	433	1	ERR2_MOUSE	Q61539 mus musculu
46	596.5	21.3	519	1	ERR1_HUMAN	P11474 homo sapien
47	584.5	20.8	462	1	ERR1_MOUSE	O08580 mus musculu
48	553.5	19.7	174	1	ESR1_BOVIN	P49884 bos taurus
49	471	16.8	467	1	RXRG_CHICK	P28701 gallus gall
50	469.5	16.7	923	1	PRGR_MOUSE	Q00175 mus musculu
51	465	16.6	924	1	PRGR_RAT	Q63449 rattus norv
52	459	16.4	458	1	PXRR_RAT	P49743 rattus norv
53	458.5	16.3	533	1	RXRB_HUMAN	P28702 homo sapien
54	457.5	16.3	758	1	GCR_ONCMY	P49843 oncorhynchu
55	456.5	16.3	377	1	PRGR_SHEEP	Q28590 ovis aries
56	456	16.3	520	1	RXRB_MOUSE	P28704 mus musculu
57	453	16.1	930	1	PRGR_RABIT	P06186 oryctolagus
58	452	16.1	463	1	RXRG_MOUSE	P28705 mus musculu
59	451.5	16.1	703	1	GCR_PIG	Q9nlu3 sus scrofa
60	451	16.1	933	1	PRGR_HUMAN	P06401 homo sapien
61	447.5	16.0	601	1	NR41_MOUSE	P12813 mus musculu
62	447.5	16.0	984	1	MCR_HUMAN	P08235 homo sapien
63	446	15.9	441	1	RXRA_BRARE	Q90416 brachydanio
64	446	15.9	463	1	RXRG_HUMAN	P48443 homo sapien
65	443.5	15.8	884	1	ANDR_EULFC	Q97776 eulemur ful
66	442.5	15.8	467	1	EXRA_MOUSE	P28700 mus musculu
67	442.5	15.8	467	1	RXRA_RAT	Q05843 rattus norv
68	442.5	15.8	598	1	NR41_HUMAN	P22736 homo sapien
69	442.5	15.8	786	1	PRGR_CHICK	P07812 gallus gall
70	441.5	15.7	776	1	GCR_XENLA	P49844 xenopus lae
71	441.5	15.7	981	1	MCR_RAT	P22199 rattus norv
72	441	15.7	470	1	RXRG_XENLA	P51129 xenopus lae
73	441	15.7	597	1	NR41_RAT	P22829 rattus norv
74	439.5	15.7	776	1	GCR_TUPGB	Q95267 tupaia glis
75	439	15.7	895	1	ANDR_MACFA	O97952 macaca fasc
76	439	15.7	895	1	ANDR_PAPHA	O97960 papio hamad
77	439	15.7	911	1	ANDR_PANTR	O97775 pan troglod
78	438.5	15.6	777	1	GCR_HUMAN	P04150 homo sapien
79	438	15.6	462	1	RXRA_HUMAN	P19793 homo sapien
80	438	15.6	899	1	ANDR_MOUSE	P19091 mus musculu
81	438	15.6	902	1	ANDR_RAT	P15207 rattus norv
82	437.5	15.6	919	1	ANDR_HUMAN	P10275 homo sapien
83	435.5	15.5	709	1	ANDR_RABIT	P49699 oryctolagus
84	435	15.5	472	1	USP_CHOPU	O76202 choristonea
85	435	15.5	777	1	GCR_AOTNA	P79686 aotus nancy
86	435	15.5	777	1	GCR_SAGOE	P79269 saguinus oe
87	435	15.5	777	1	GCR_SAIBB	O13186 saimiri bol
88	435	15.5	807	1	GCR_PAROL	O73673 paralichthy
89	435	15.5	907	1	ANDR_CANFA	Q9tt90 canis famil
90	432.5	15.4	598	1	NR41_CANFA	P51666 canis famil
91	431.5	15.4	778	1	GCR_SAISC	O46567 saimiri sci
92	431	15.4	461	1	USP_MANSE	P54779 manduca sex
93	427.5	15.2	577	1	NR42_XENLA	Q04913 xenopus lae
94	425.5	15.2	462	1	USP_BOMMO	P43700 bombyx mori
95	423.5	15.1	777	1	GCR_CAVPO	P49115 cavia porce
96	423.5	15.1	783	1	GCR_MOUSE	P06537 mus musculu
97	422	15.0	488	1	RXRA_XENLA	P51128 xenopus lae
98	417.5	14.9	795	1	GCR_RAT	P06536 rattus norv
99	415	14.8	612	1	MCR_XENLA	Q91573 xenopus lae
100	414	14.8	977	1	MCR_TUPGB	Q29131 tupaia glis
101	410	14.6	385	1	NR21_MOUSE	Q64104 mus musculu
102	409	14.6	385	1	NR21_HUMAN	O95466 homo sapien
103	407	14.5	385	1	NR21_CHICK	Q91379 gallus gall
104	403.5	14.4	508	1	USP_DROME	P20153 drosophila
105	397.5	14.2	495	1	NR61_MOUSE	O54249 mus musculu
106	397	14.2	598	1	NR42_HUMAN	P43354 homo sapien

107	396.5	14.1	598	1	NR42_RAT	Q07917 rattus norv
108	394.5	14.1	598	1	NR42_MOUSE	Q06219 mus musculu
109	391.5	14.0	486	1	NR21_XENLA	P70052 xenopus lae
110	385.5	13.7	458	1	RRA_XENLA	P51126 xenopus lae
111	384.5	13.7	808	1	FTFB_DROME	Q05192 drosophila
112	384	13.7	480	1	NR61_HUMAN	Q15406 homo sapien
113	384	13.7	560	1	NR52_MOUSE	P45448 mus musculu
114	383	13.7	396	1	NR21_ORYLA	Q9yq13 oryzias lat
115	383	13.7	541	1	NR52_HUMAN	O00482 homo sapien
116	382	13.6	447	1	RRA_FUGRU	Q9w523 fugu rubrip
117	382	13.6	501	1	NR52_CHICK	Q42101 gallus gall
118	377.5	13.5	628	1	NR43_RAT	P51179 rattus norv
119	376	13.4	462	1	RRA_MOUSE	P11416 mus musculu
120	376	13.4	675	1	ECR_AEDAE	P49880 aedes aegyp

ALIGNMENTS

```
RESULT 1
ESR2_HUMAN
ID   ESR2_HUMAN      STANDARD;      PRT;   530 AA.
AC   Q92731; O60685; O60608; O75584; O60702; O60703; O75583; Q9UQK9;
AC   Q9UEV6; Q9UHD3;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2 OR ESTRB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM BETA-1).
RX   MEDLINE=98139878; PubMed=9473491;
RA   Ogawa S., Inoue S., Watanabe T., Hiroi H., Orimo A., Hosoi T.,
RA   Ouchi Y., Muramatsu M.;
RT   "The complete primary structure of human estrogen receptor beta
RT   (hERbeta) and its heterodimerization with ER alpha in vivo and in
RT   vitro.";
RL   Biochem. Biophys. Res. Commun. 243:122-126(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Testis, and Breast;
RX   MEDLINE=98300286; PubMed=9636657;
RA   Moore J.T., McKee D.D., Slentz-Kesler K., Moore L.B., Jones S.A.,
RA   Horne E.L., Su J.-L., Kliewer S.A., Lehmann J.M., Willson T.M.;
RT   "Cloning and characterization of human estrogen receptor beta
RT   isoforms.";
RL   Biochem. Biophys. Res. Commun. 247:75-78(1998).
RN   [3]
RP   SEQUENCE OF 48-540 FROM N.A. (ISOFORM BETA 1), AND CHARACTERIZATION.
RC   TISSUE=Testis;
RX   MEDLINE=96354875; PubMed=8769318;
RA   Mosselman S., Polman J., Dijkema R.;
RT   "ER beta: identification and characterization of a novel human
RT   estrogen receptor.";
RL   FEBS Lett. 392:49-53(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM BETA 2/CX), AND CHARACTERIZATION.
RC   TISSUE=Testis;
RX   MEDLINE=98347908; PubMed=9671811;
RA   Ogawa S., Inoue S., Watanabe T., Orimo A., Hosoi T., Ouchi Y.,
RA   Muramatsu M.;
RT   "Molecular cloning and characterization of human estrogen receptor
RT   beta cx: a potential inhibitor of estrogen action in human.";
RL   Nucleic Acids Res. 26:3505-3512(1998).
RN   [5]
RP   SEQUENCE OF 59-540 FROM N.A. (ISOFORMS BETA-5A AND BETA-6).
RC   TISSUE=Endometrium;
RA   Brandenberger A.W., Lebovic D., Taylor R.N., Jaffe R.B.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM BETA-2A).
RC   TISSUE=Ovary;
RX   MEDLINE=98348389; PubMed=9685228;
RA   Lu B., Leygue E., Dotzlaw H., Murphy L.J., Murphy L.C., Watson P.H.;
RT   "Estrogen receptor-beta mRNA variants in human and murine tissues.";
RL   Mol. Cell. Endocrinol. 138:199-203(1998).
RN   [7]
RP   SEQUENCE OF 1-69 FROM N.A.
RA   Li L.C., Dahiya R.;
RT   "Cloning and characterization of the estrogen receptor beta gene
RT   promoter.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   CHARACTERIZATION.
RX   MEDLINE=97467383; PubMed=9325313;
RA   Pace P., Taylor J., Suntharalingam S., Coombes R.C., Ali S.;
RT   "Human estrogen receptor beta binds DNA in a manner similar to and
RT   dimerizes with estrogen receptor alpha.";
RL   J. Biol. Chem. 272:25832-25838(1997).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF ER
CC       ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN DEPENDENT MANNER.
CC       ISOFORM BETA CX LACKS LIGAND BINDING ABILITY AND HAS NO OR ONLY
CC       VERY LOW ERE BINDING ACTIVITY RESULTING IN THE LOSS OF LIGAND-
CC       DEPENDENT TRANSACTIVATION ABILITY. DNA-BINDING BY ER-ALPHA AND
CC       BETA IS RAPIDLY LOST AT 37 DEGREES CELSIUS IN THE ABSENCE OF
CC       LIGAND WHILE IN THE PRESENCE OF 17 BETA-ESTRADIOL AND 4 HYDROXY-
CC       TAMOXIFEN LOSS IN DNA-BINDING AT ELEVATED TEMPERATURE IS MORE
CC       GRADUAL.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER
CC       ALPHA. ISOFORM BETA-2/CX PREFERENTIALLY FORMS A HETEREODIMER WITH
CC       ER-ALPHA RATHER THAN ER-BETA AND INHIBITS DNA-BINDING BY ER-ALPHA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 8 ISOFORMS, BETA 1 (SHOWN HERE), BETA 2/CX,
CC       BETA-2A, BETA-3, BETA-4, BETA-5, BETA-5A AND BETA-6; ARE PRODUCED
CC       BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM BETA-1 IS EXPRESSED IN TESTIS AND
CC       OVARY, AND AT A LOWER LEVEL IN HEART, BRAIN, PLACENTA, LIVER,
CC       SKELETAL MUSCLE, SPLEEN, THYMUS, PROSTATE, COLON, BONE MARROW,
CC       MAMMARY GLAND AND UTERUS. ALSO FOUND IN UTERINE BONE, BREAST, AND
CC       OVARIAN TUMOR CELL LINES, BUT NOT IN COLON AND LIVER TUMORS.
CC       ISOFORM BETA 2 IS EXPRESSED IN SPLEEN, THYMUS, TESTIS AND OVARY
CC       AND AT A LOWER LEVEL IN SKELETAL MUSCLE, PROSTATE, COLON, SMALL
CC       INTESTINE, LEUKOCYTES, BONE MARROW, MAMMARY GLAND AND UTERUS.
CC       ISOFORM BETA-3 IS FOUND IN TESTIS. ISOFORM BETA 4 IS EXPRESSED IN
CC       TESTIS, AND AT A LOWER LEVEL IN SPLEEN, THYMUS, OVARY, MAMMARY
CC       GLAND AND UTERUS. ISOFORM BETA-5 IS EXPRESSED IN TESTIS, PLACENTA,
CC       SKELETAL MUSCLE, SPLEEN AND LEUKOCYTES, AND AT A LOWER LEVEL IN
CC       HEART, LUNG, LIVER, KIDNEY, PANCREAS, THYMUS, PROSTATE, COLON,
CC       SMALL INTESTINE, BONE MARROW, MAMMARY GLAND AND UTERUS. NOT
CC       EXPRESSED IN BRAIN.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   --  ------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AB006590; BAA24953.1; -.
DR   EMBL; AF051427; AAC05985.1; -.
DR   EMBL; AF051428; AAC05751.1; -.
DR   EMBL; AF061054; AAC39784.1; -.
DR   EMBL; AF061055; AAC39785.1; -.
DR   EMBL; AF060555; AAC15234.1; -.
DR   EMBL; X99101; CAA67555.1; ALT_INIT.
DR   EMBL; AB006589; BAA31966.1; -.
```

```
DR   EMBL; AF074598; AAC25602.1; -.
DR   EMBL; AF074599; AAC25603.1; -.
DR   EMBL; AF124790; AAD32580.1; -.
DR   EMBL; AF047463; AAC03786.1; -.
DR   EMBL; AF191544; AAF24232.1; -.
DR   TRANSFAC; T04651; -.
DR   Genew; HGNC:3468; ESR2.
DR   MIM; 601663; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA binding; Nuclear protein;
KW   Zinc finger; Steroid binding; Alternative splicing; Phosphorylation.
FT   DOMAIN        1    148       MODULATING.
FT   DNA_BIND    149    214       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     149    169       C4-TYPE.
FT   ZN_FING     185    209       C4-TYPE.
FT   DOMAIN      215    530       STEROID-BINDING.
FT   MOD_RES      87     87       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     488    488       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC    319    323       EVELS -> MRGNA (IN ISOFORM BETA-2A).
FT   VARSPLIC    319    409       MISSING (IN ISOFORM BETA-5A).
FT   VARSPLIC    324    530       MISSING (IN ISOFORM BETA-2A).
FT   VARSPLIC    365    375       DRGKCVEGILE -> YVPSGHSDPGC (IN ISOFORM
FT                                BETA-6).
FT   VARSPLIC    376    530       MISSING (IN ISOFORM BETA-6).
FT   VARSPLIC    469    530       SNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITG
FT                                SECSPAEDSKSKEGSQNPQSQ -> RAEKASQTITSFGMKM
FT                                ETLLPEATMEQ (IN ISOFORM BETA-2/2X).
FT   VARSPLIC    469    530       SNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITG
FT                                SECSPAEDSKSKEGSQNPQSQ -> SSLSISWRIFMLREAS
FT                                CHGVRQTPGGAHMSVSPRSPSEAGQQPRE (IN ISOFORM
FT                                BETA-3).
FT   VARSPLIC    469    530       SNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITG
FT                                SECSPAEDSKSKEGSQNPQSQ -> RWGEKQFIHIKLS
FT                                (IN ISOFORM BETA-4).
FT   VARSPLIC    469    530       SNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITG
FT                                SECSPAEDSKSKEGSQNPQSQ -> RYAP (IN ISOFORM
FT                                BETA-5).
FT   CONFLICT     48     48       F -> G (IN REF. 3).
SQ   SEQUENCE   530 AA;  59216 MW;  8CAE342159992418A CRC64;

  Query Match           100.0%;  Score 2805;  DB 1;  Length 530;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-218;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240

Qy        241 CAGKAKRSGCHAPVRPELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGKAKRSGCHAPVRPELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530


RESULT 2
ESR2_CALJA
ID   ESR2_CALJA     STANDARD;      PRT:   530 AA.
AC   Q95171;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC   Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Gaughan J., Scobie G.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------- ----------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------- ----------------------------------------
DR   EMBL; Y09372; CAA70546.2; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid binding; Phosphorylation.
FT   DOMAIN        1    148       MODULATING.
FT   DNA_BIND    149    214       NUCLEAR RECEPTOR-TYPE.
```

```
FT   ZN_FING      149     169        C4 TYPE.
FT   ZN_FING      185     209        C4 TYPE.
FT   DOMAIN       215     530        STEROID BINDING.
FT   MOD_RES       87      87        PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES      488     488        PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   530 AA;  59087 MW;  45D89107A84C54D1 CRC64;

  Query Match             93.2%;  Score 2615;  DB 1;  Length 530;
  Best Local Similarity   92.4%;  Pred. No. 8.6e-203;
  Matches  489;  Conservative   20;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||  |||||||| ||||||||:|||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNFGQSILPLEHGPIYIPSSYVESHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              ||||| |||| |||:|| |||||||| || ||||||||||||||||||||||||||||:
Db         61 SVTNLEEGPGRQTTSPNMLWSTPGHLSPLAVHHQLSHLYAEPQKSPWCEARSLEHTLPVS 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKVSGNHCASPVTGPSSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| :|:||||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQGNAEEQLR 240

Qy        241 CAGKAKRSGGHAPRVRELLLDAISPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |||||||||| |||||||| ||||||||||||||||||||||| ||||||||||||||
Db        241 CAGKAKRSGGHVPRVRELLLSAISPEQLVLTLLEAEPPHVLISEPSVPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||:|||||||||||||||||| |||||||||:||||:|||||||||||||||||||:|
Db        301 LADEELVHMISWAKKIPGFVELSLLDQVPLLESCWLEVLMVGLMWRSIDHPGKLIFAPNL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              :|||||||||||||:|||||||||||||||||||||||||:||||   |||||||||
Db        361 ILDRDEGKCVEGILEVFDMLLATTSRFRELKLQHKEYLCVKAMVLLNSQVDPLVTATQDA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :|||:|||||||||||||||||| |||||:|||||||||||:|||||||||||:||||||
Db        421 ESSQKLAHLLNAVTDALVWVIAKSGFSSQQQSVRLANLLMLLSHIRHASNKGMEHLLSMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQS 529
              |||||||||||||:|||:|||||||||||||||||||| |||||||
Db        481 CKNVVPVYDLLLEMMNAHVVRGCKSSITGSECSPAEDSKSTEGSQNPQS 529


RESULT 4
ESR2_RAT
ID   ESR2_RAT       STANDARD;      PRT;   530 AA.
AC   Q62986; Q70195; O35785; O55015; O55016; O35784; Q9R185;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER beta).
GN   ESR2 OR NR3A2 OR ERBETA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 46-530 FROM N.A. (ISOFORM BETA-1).
RC   STRAIN=Sprague-Dawley; TISSUE=Prostate;
RX   MEDLINE=96234066; PubMed=8650195;
RA   Kuiper G.G.J.M., Enmark E., Pelto-Huikko M., Nilsson S.,
RA   Gustafsson J.-A.;
RT   "Cloning of a novel receptor expressed in rat prostate and ovary.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:5925-5930(1996).
RN   [2]
RP   SEQUENCE OF 46-530 FROM N.A. (ISOFORM BETA-2).
RC   STRAIN=Wistar; TISSUE=Ovary;
RX   MEDLINE=98262932; PubMed=9600083;
RA   Maruyama K., Endoh H., Sasaki-Iwaoka H., Kanou H., Shimaya E.,
RA   Hashimoto S., Kato S., Kawashima H.;
RT   "A novel isoform of rat estrogen receptor beta with 18 amino acid
RT   insertion in the ligand binding domain as a putative dominant
RT   negative regular of estrogen action.";
RL   Biochem. Biophys. Res. Commun. 246:142-147(1998).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM BETA-2).
RC   STRAIN=Wistar; TISSUE=Prostate;
RA   Aldridge T.C.;
RT   "Tissue specific responses to estrogen: an explanation based on
RT   differential activation of multiple estrogen receptors with different
RT   estrogen response elements.";
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 46-530 FROM N.A. (ISOFORMS BETA-1; BETA-2; DELTA-3).
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=98151005; PubMed=9492041;
RA   Petersen D.N., Tkalcevic G.T., Koza-Taylor P.H., Turi T.G.,
RA   Brown T.A.;
RT   "Identification of estrogen receptor beta2, a functional variant of
RT   estrogen receptor beta expressed in normal rat tissues.";
RL   Endocrinology 139:1082-1092(1998).
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM BETA-1/DELTA-4).
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RA   Price R., Handa R.J.;
RT   "A novel splice variant of estrogen receptor beta found in rat
RT   brain.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF ER-
CC       ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC       ISOFORMS BETA-1/DELTA-3 AND BETA-2/DELTA-3 ARE UNABLE TO BIND DNA
CC       AND ACTIVATE TRANSCRIPTION DUE TO THE TRUNCATION OF THE DNA
CC       BINDING DOMAIN. ISOFORM BETA-2 SHOWS LOSS OF LIGAND BINDING
CC       AFFINITY AND SUPPRESSES ER-ALPHA AND ER-BETA-1 MEDIATED
CC       TRANSCRIPTIONAL ACTIVATION AND MAY ACT AS A DOMINANT NEGATIVE
CC       REGULATOR OF ESTROGEN ACTION.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA. CAN ALSO FORM HETERODIMERS BETWEEN ISOFORMS BETA-1 AND
CC       BETA-2.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; BETA-1 (SHOWN HERE),
CC       BETA-1/DELTA-3, BETA-1/DELTA-4, BETA-2 AND BETA-2/DELTA-3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN PROSTATE, OVARY, LUNG, LIVER,
CC       KIDNEY, FAT, BONE, BRAIN, UTERUS AND TESTIS.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U57439; AAC52602.1; -.
DR   EMBL; AB012721; BAA25431.1; -.
DR   EMBL; AJ002602; CAA05631.1; ALT_INIT.
DR   EMBL; AJ002603; CAA05632.1; -.
DR   EMBL; AF042058; AAB97424.1; -.
DR   EMBL; AF042059; AAB97425.1; -.
DR   EMBL; AF042060; AAB97426.1; -.
DR   EMBL; AF042061; AAB97427.1; -.
DR   EMBL; AF161187; AAD47637.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
```

```
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding; Alternative splicing; Phosphorylation.
FT   DOMAIN        1    148       MODULATING.
FT   DNA_BIND    149    214       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     149    169       C4-TYPE.
FT   ZN_FING     185    209       C4-TYPE.
FT   DOMAIN      215    530       STEROID-BINDING.
FT   MOD_RES      87     87       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     488    488       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC    179    217       MISSING (IN ISOFORM BETA-1/DELTA-3 AND
FT                                ISOFORM BETA-2/DELTA-3).
FT   VARSPLIC    219    318       MISSING (IN ISOFORM BETA-1/DELTA-4).
FT   VARSPLIC    364    464       P -> PSSEDPHWHVAQMKSAAPR (IN ISOFORM
FT                                BETA-2 AND ISOFORM BETA-2/DELTA-3).
FT   CONFLICT     72     72       L -> Q (IN REF. 3 AND 4)
FT   CONFLICT    150    150       P -> A (IN REF. 3 AND 4)
FT   CONFLICT    165    165       S -> P (IN REF. 3; CAA05631).
FT   CONFLICT    505    505       S -> P (IN REF. 3; CAA05631).
SQ   SEQUENCE   530 AA;  59152 MW;  36F269D9FD773DA9 CRC64;

  Query Match          89.0%;  Score 2496;  DB 1;  Length 530;
  Best Local Similarity   88.3%;  Pred. No. 3.2e-193;
  Matches  468;  Conservative   27;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:|||||||:||:|||||||||||||| |||||||| ||| |||||||||||:|
Db          1 MEIKNSPSSLSSPASYNCSQSILPLEHGPIYIPSSYVDNRHEYSAMTFYSPAVMNYSVPG 60

Qy         61 NVTNLEGGPGRQTTSPNVIWPTPGHISPLVVHPQLSHLYAEPQKSPWCEAKSLEHTLPVN 120
              : :||:||| | :|||||||| |||||  | | | ||||||||||||||||||||||
Db         61 STSNLDGGPVRLSTSPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEAKSLEHTLPVN 120

Qy        121 RPTIKPKVSQNPCASPVTCPDSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||:|| |||||| | :||||||| ||||||||||||||||||||||||||||||
Db        121 RPTIKPKLSGSSCASPVTSPNAKRDAHFCPVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSERERCGYRLVRRQRSAEDIH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||| |:|
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSERERCGYRLVRRQRSSERQVH 240

Qy        241 CAGKAKRSGGHAPRVKEILISTLSPPQIVGTLLEAEPPHVLISPPSAPFTEASMMMSLTK 300
              | ||||:||||||:|||| |||||||||||||||:||:|||| ||||||||||||
Db        241 CLSKAKRNGGHAPRVKEILISTLSPPQIVGTLLEAEPPNVLVSRPSMPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||| :| |:|
Db        361 VLDRDEGKCVEGILEIFDMILATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEA 420

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :|||| ||||||||||||||||||||||||||:|||||||||||| ||||||||:||
Db        421 FSSRFLTHLLNAVTDALVWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              |||||||||||||||| ||| |||:|||||  ||||:|| ||| |||
Db        481 CKNVVPVYDLLLEMLNAHTLRGYKSSISGSECSSTEDSKNKESSQNLQSQ 530


RESULT 4
ESR2_MOUSE
ID   ESR2_MOUSE     STANDARD;      PRT;   530 AA.
AC   O08537; O35635; O70519;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Estrogen receptor beta (ER beta).
GN   ESR2 OR NR3A2 OR ESTRB.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 46-530 FROM N.A., AND MUTAGENESIS.
RC   STRAIN=129/Sv; TISSUE=Ovary;
RX   MEDLINE=97211383; PubMed=9058381;
RA   Tremblay G.B., Tremblay A., Copeland N.G., Gilbert D.J., Jenkins N.A.,
RA   Labrie F., Giguere V.;
RT   "Cloning, chromosomal localization, and functional analysis of the
RT   murine estrogen receptor beta.";
RL   Mol. Endocrinol. 11:353-365(1997).
RN   [2]
RP   SEQUENCE OF 46-530 FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Ovary;
RX   MEDLINE=97424142; PubMed=9280064;
RA   Pettersson K., Grandien K., Kuiper G.G.J.M., Gustafsson J.-A.;
RT   "Mouse estrogen receptor beta forms estrogen response element-binding
RT   heterodimers with estrogen receptor alpha.";
RL   Mol. Endocrinol. 11:1486-1496(1997).
RN   [3]
RP   SEQUENCE OF 1-60 FROM N.A.
RC   TISSUE=Ovary;
RA   Leyque E., Lu B., Dotzlaw H., Glor C., Watson P.H., Murphy L.C.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-60 FROM N.A.
RC   STRAIN=C57BL/6J X 129; TISSUE=Ovary;
RA   Rosenfeld C.S., Lubahn D.B.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   CARBOHYDRATE-LINKAGE SITE SER-61, AND PHOSPHORYLATION SITE SER-61.
RX   MEDLINE=20450734; PubMed=10995228;
RA   Cheng X., Cole R.N., Zaia J., Hart G.W.;
RT   "Alternative O-glycosylation/O-phosphorylation of the murine estrogen
RT   receptor beta.";
RL   Biochemistry 39:11609-11620(2000).
RN   [6]
RP   ALTERNATIVE SPLICING.
RC   TISSUE=Ovary;
RX   MEDLINE=98348389; PubMed=9685228;
RA   Lu B., Leyque E., Dotzlaw H., Murphy L.J., Murphy L.C., Watson P.H.;
RT   "Estrogen receptor-beta mRNA variants in human and murine tissues.";
RL   Mol. Cell. Endocrinol. 138:199-203(1998).
RN   [7]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=98268825; PubMed=9607809;
RA   Rosenfeld C.S., Ganjam V.K., Taylor J.A., Yuan X., Stiehr J.R.,
RA   Hardy M.P., Lubahn D.B.;
RT   "Transcription and translation of estrogen receptor-beta in the male
RT   reproductive tract of estrogen receptor-alpha knock-out and wild-type
RT   mice.";
RL   Endocrinology 139:2982-2987(1998).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; BETA-1 (SHOWN HERE),
CC       BETA-2, BETA-5, BETA-5A AND BETA-6; ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN PROSTATE, OVARY, LEYDIG CELLS AND
CC       IN EPITHELIUM OF THE EFFERENT DUCTULES AND OF THE INITIAL SEGMENT
CC       OF THE EPIDIDYMIS.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N TERMINAL DOMAIN,
```

```
CC       A DNA BINDING DOMAIN AND A C TERMINAL STEROID BINDING DOMAIN.
CC   !   SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; U81451; AAB51132.1; -.
DR   EMBL; AJ000220; CAA03949.1; -.
DR   EMBL; AF067422; AAC17919.1; ALT_INIT.
DR   EMBL; AF063853; AAC16656.1; -.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T04649; -.
DR   GlycoSuiteDB; 008537; -.
DR   MGD; MGI:109392; Esr2.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid binding; Alternative splicing; Glycoprotein;
KW   Phosphorylation.
FT   DOMAIN        1    148       MODULATING.
FT   DNA_BIND    149    214       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     149    169       C4-TYPE.
FT   ZN_FING     185    209       C4-TYPE.
FT   DOMAIN      215    530       STEROID BINDING.
FT   MOD_RES      61     61       PHOSPHORYLATION.
FT   MOD_RES      87     87       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     105    105       PHOSPHORYLATION (BY MAPK) (PROBABLE).
FT   MOD_RES     488    488       PHOSPHORYLATION (BY SIMILARITY).
FT   CARBOHYD     61     61       O-LINKED (GLCNAC).
FT                                /FTId=CAR_000201.
FT   VARSPLIC    364    364       R   -> RSSEDPHWHVAQTKSAVPR (IN ISOFORM
FT                                BETA-2).
FT   VARSPLIC    319    364       MISSING (IN ISOFORM BETA-5).
FT   VARSPLIC    365    409       MISSING (IN ISOFORM BETA-6).
FT   VARSPLIC    319    409       MISSING (IN ISOFORM BETA-5A).
FT   MUTAGEN     105    105       S->A: ABOLISHES STIMULATORY EFFECT OF
FT                                RAS.
FT   MUTAGEN     139    139       S->A: NO LOSS OF THE STIMULATORY EFFECT
FT                                OF RAS.
FT   CONFLICT     47     47       T -> A (IN REF. 2).
FT   CONFLICT    142    142       A   -> T (IN REF. 2).
FT   CONFLICT    200    200       S   -> N (IN REF. 2).
FT   CONFLICT    378    378       G   -> D (IN REF. 2).
FT   CONFLICT    412    412       P   -> H (IN REF. 2).
FT   CONFLICT    445    445       G   -> R (IN REF. 2).
FT   CONFLICT    511    511       E   -> G (IN REF. 2).
SQ   SEQUENCE   530 AA;  59012 MW;  90E73A161550ABE4 CRC64;

  Query Match             88.9%;  Score 2495;  DB 1;  Length 530;
  Best Local Similarity   88.5%;  Pred. No. 3.9e-193;
  Matches  469;  Conservative   25;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:|||||||| ||:|||||||||||||| |||||||||:| ||| ||||||||||||:||
Db          1 MEIKNSPSSLTSPASYNCSQSILPLEHGPIYIPSSYVESRHEYSAMTFYSPAVMNYSVPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              :  |||||| ||| |||||||| |||||  | | | ||||||||||||||||||||||||
```

```
Db         61 STGNLEGGFVRQTASPNVLWFTSGHLSPLATHCQSSLLYAEPQKSPWCEAKSLEHTLFVN 120

Qy        121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||| |: |||||| | :|||||||||||||||||||||||||||||||||||||||
Db        121 RETLKRKLGGSGCASPVTSPSAKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||:||||||| |||
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSASEQVH 240

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  |||| || |||:||||:|||||||||||||||:||:|||| ||||||||||||
Db        241 CLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||||||| |||||||||||:|||||||||||||||||||||||||
Db        301 LADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              |||||||||||||||| ||||||:|||||||||||||||||||||||||||| ||:|:|
Db        361 VLDRDEGKCVEGILEIFGMLLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEA 420

Qy        421 ESSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :|||||  |||||||||||||||:|||||||||:|||||||||||||| ||||||||:||
Db        421 ESSRKLTHLLNAVTDALVWVISKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMK 480

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              |||||||||||||||| ||| |||:|||||  ||||||||||| |||
Db        481 CKNVVPVYDLLLEMLNAHILRGYKSSISGSECSSAEDSKSSKEGSQNLQSQ 530


RESULT 5
ESR2_BOVIN
ID   ESR2_BOVIN     STANDARD;      PRT;   527 AA.
AC   Q9XSB5; Q9TTS2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Holstein; TISSUE=Ovarian follicle;
RX   MEDLINE=99150196; PubMed=10026117;
RA   Rosenfeld C.S., Yuan X., Manikkam M., Calder M.D., Garverick H.A.,
RA   Lubahn D.B.;
RT   "Cloning, sequencing, and localization of bovine estrogen receptor
RT   beta within the ovarian follicle.";
RL   Biol. Reprod. 60:691-697(1999).
RN   [2]
RP   SEQUENCE OF 4-527 FROM N.A.
RX   MEDLINE=99359179; PubMed=10472221;
RA   Walther N., Hoogtas C., Tillmann G., Ivell R.;
RT   "Cloning of bovine estrogen receptor beta (Erbeta): expression of
RT   novel deleted isoforms in reproductive tissues.";
RL   Mol. Cell. Endocrinol. 152:37-45(1999).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC       MAY PLAY A ROLE IN OVARIAN FOLLICULAR GROWTH AND MATURATION.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: PRESENT IN GRANULOSA CELLS OF ANTRAL FOLLICLES
CC       IN VARIOUS STAGES OF FOLLICULAR GROWTH.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
```

```
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF110402; AAD24432.1; -.
DR   EMBL; Y18017; CAB53861.1; ALT_INIT.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding; Phosphorylation.
FT   DOMAIN        1    145       MODULATING.
FT   DNA_BIND    146    211       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     146    166       C4 TYPE.
FT   ZN_FING     182    206       C4 TYPE.
FT   DOMAIN      212    527       STEROID-BINDING.
FT   MOD_RES      84     84       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     485    485       PHOSPHORYLATION (BY SIMILARITY).
FT   CONFLICT     60     60       N -> D (IN REF. 2).
SQ   SEQUENCE   527 AA;  59031 MW;  9CEFFE106F4E4C84 CRC64;

  Query Match             87.8%;  Score 2463.5;  DB 1;  Length 527;
  Best Local Similarity   87.9%;  Pred. No. 1.3e-190;
  Matches  466;  Conservative   22;  Mismatches   39;  Indels    3;  Gaps    1;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||:|||||||||| |||| ||||||| | ||:||||:| ||| |:|||||||||||:
Db          1 MDVKNSPSSLNSPVSYNCGQSILPLEPGPIYLPSSYVESRHEYSAVTFYSPAVMNYSIPN 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
                | | |||||||||||||||||||||||| :| | | |||||||||| | ||||||||||
Db         61 ---NSEDGPGRQTTSPNVLWPTPGHLSPLAIHCQPSVLYAEPQKSPWRETRSLEHTLPVN 117

Qy        121 PETLKRKVSGNPCASPVTGPQSKPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||| ||: |||| | | ||||||||||||||||||||||||||||||||||||||
Db        118 RETLKRKASGSSCASPATSPSSKPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 177

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRMLKHKRQRDDGEG 240
              ||||||||||||||||||||||||||||||||||||||||||||||:||||:|||||||
Db        178 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRMLKHKRQRDDGEG 237

Qy        241 CAGKAKRSGGHAPRVRELLIDALSPFQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              | | ||:||   ||:|||| ||||||||||||||||||||||||| ||||||||||||
Db        238 CLSKTKRNGGPMTEVKELIGALSPFQLVLTLLEAEPPHVLISRPSTPFTEASMMMSLTK 297

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||:||||||||:||||:|||||||||||||||||||||||
Db        298 LADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLEVLMVGLMWRSIEHPGKLIFAPDL 357

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              :||||||||||||||||||||||||||||||||||||||||||||||||  || ||:|
Db        358 ILDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPSATAPQEA 417

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              || ||| |||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db        418 DSGRKLTHLLNAVTDALVWVIAKSGMSSQQQSMRLANLLMLLSHVPHASNKGMEHLLNMK 477

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              |||||||||||||||| ||| || :|||| :  |||:|||||| ||:|
Db        478 CKNVVPVYDLLIEMLNAHILRGNKSIVTGSERNLVEDSESKEGSQKPQAQ 527
```

```
RESULT 6
ESR2_SHEEP
ID   ESR2_SHEEP     STANDARD;      PRT;   527 AA.
AC   Q9TU15; Q9N0T6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS BETA AND BETA-1).
RC   TISSUE=Ovary;
RX   MEDLINE=21313342; PubMed=11420232;
RA   Cardenas H., Burke K.A., Bilsby R.M., Pope W.F., Nephew K.P.;
RT   "Estrogen receptor beta in the sheep ovary during the estrous cycle
RT   and early pregnancy.";
RL   Biol. Reprod. 65:128-134(2001).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; BETA (SHOWN HERE) AND
CC       BETA-1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF177936; AAD55772.1; -.
DR   EMBL; AF257109; AAF71745.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding; Alternative splicing; Phosphorylation.
FT   DOMAIN        1    145       MODULATING.
FT   DNA_BIND    146    211       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     146    166       C4-TYPE.
FT   ZN_FING     182    206       C4-TYPE.
FT   DOMAIN      212    527       STEROID-BINDING.
FT   MOD_RES      84     84       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     485    485       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC    316    324       FVELSLYDQ -> MKGNVLKEF (IN ISOFORM BETA-
FT                                1).
FT   VARSPLIC    325    527       MISSING (IN ISOFORM BETA-1).
SQ   SEQUENCE   527 AA;  59120 MW;  9CD7A3E89497EDEA CRC64;
```

```
  Query Match             87.5%;  Score 2455.5;  DB 1;  Length 527;
  Best Local Similarity   87.5%;  Pred. No. 5.8e-190;
  Matches  464;  Conservative   24;  Mismatches   39;  Indels    3;  Gaps    1;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||:|||||||||| |||| ||||||| | ||:|||||:| ||| |:|||||||||||||:
Db          1 MDVKNSPSSLNSPVSYNCGQSILPLEPGPIYLPSSYVESRHEYSAVTFYSPAVMNYSIPN 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              | | |||||||||||||||||||||| :| | | |||||||||||| |||||| |||
Db         61 --NSEDGPGRQTTSPNVLWPTPGHLSPLAIHCQSSLLYAEPQKSPWCETRSLEHTFPVN 117

Qy        121 RLTLKRKVSGNKCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||| ||: |||||: | ||||||||||||||||||||||||||||||||||||||
Db        118 RETLKRKASGSSCASPVSSPSSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 177

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||:||||::||||
Db        178 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRIVRRQRNSDEQLH 237

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              | | ||:|    ||:|||| ||||||||||||||||||||:|||||||||||||||||
Db        238 CLSKTKRNGAPMTRVKELLLSALSPEQLVLTLLEAEPPHVLMSRPSAPFTEASMMMSLTK 297

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||:||||||||:||||:|||||||||||||||||||||||||
Db        298 LADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLEVLMVGLMWRSIDHPGKLIFAPDL 357

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||  ||:|:|
Db        358 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPSATASQEA 417

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              || ||| |||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        418 DSGRKLTHLLNAVTDALVWVIAKSGMSSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 477

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
              |||||||||||||||| ||  |  :| || : ||||:|||||| ||:|
Db        478 CKNVVPVYDLLLEMLNAHTLRSNKPLVTRSERNLAEDSESKEGSQKPQAQ 527
```

```
RESULT 7
ESR2_PIG
ID   ESR2_PIG        STANDARD;       PRT;   526 AA.
AC   Q9XSW2; Q98UW5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Kowalski A.A., Graddy L.G., Vale Cruz D.S., Choi I.,
RA   Katzenellenbogen B.S., Simmen F.A., Simmen R.C.M.;
RT   "Molecular cloning of porcine estrogen receptor beta cDNAs and
RT   developmental expression in peri-implantation embryos.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   LaVoie H.A., DeSimone D.C.;
RT   "Cloning and expression of estrogen receptor beta isoforms from
RT   porcine ovary.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF164957; AAD45381.2; -.
DR   EMBL; AF267746; AAK15151.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid binding; Phosphorylation.
FT   DOMAIN          1    145       MODULATING.
FT   DNA_BIND      146    211       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING       146    166       C4-TYPE.
FT   ZN_FING       182    206       C4-TYPE.
FT   DOMAIN        212    526       STEROID-BINDING (BY SIMILARITY).
FT   MOD_RES        84     84       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES       484    484       PHOSPHORYLATION (BY SIMILARITY).
FT   CONFLICT      317    317       M -> V (IN REF. 2).
FT   CONFLICT      469    469       T -> M (IN REF. 2).
SQ   SEQUENCE   526 AA;  58849 MW;  45CE2BF66107BBF6 CRC64;
```

```
  Query Match             86.1%;  Score 2415;  DB 1;  Length 526;
  Best Local Similarity   86.2%;  Pred. No. 1e-186;
  Matches  457;  Conservative   29;  Mismatches   40;  Indels    4;  Gaps    2;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |||||||:|||| |||||||:|||| | ||||||||:| ||| ||||||||:|||| |
Db          1 MDIKNSPSNLNSPVSYNCSQSVLPLEPGPIYIPSSYVESCHEYSAMTFYSPAVVNYSISS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
                | | |||| |||||||||||||| :| | | ||||||||||| ||||||||||
Db         61 ---NSEVGPGRQATSPNVLWPTPGHLSPIAIHCQPSLIYAEPQKSPWCDTRSLEHTLPVN 117

Qy        121 RETLKRKVSGNKCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              ||||||| ||: |||||: | ||||||||||||||||||||||||||||||||||||||
Db        118 RETLKRKASGSSCASPVISPSSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 177

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRLVRRQRSADEQLH 240
              |||||||||||||||||||||||||||||||||||||||||||||:||:|||:  ||
Db        178 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRREKCGYRIVRKQRNSEGHLH 237

Qy        241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              |  :|:::( |  ||:|||  ||||||||||||||||||:|||| ||||||||||||
Db        238 CLSKAKKNGGHTTRVKELLLSILSPEQLVLTLLEAEPPHVLVSRPSTPFTEASMMMSLTK 297

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||:||||:||||||||:||||:|||||||||||||||||||||
Db        298 LADKELVHMISWAKKIPGFMELSLYDQVRLLESCWLEVLMVGLMWRSIDHPGKLIFAPDL 357

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||  | |:|
Db        358 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYP-SAAAQEA 416

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :|||| ||||||||||||||:|||||||:|||||||||||||||||||| |||||||
Db        417 ESSRKLTHLLNAVTDALVWVIARSGISSQQQSVRLANLLMLLSHVRHASNKGTEHLLNMK 476

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ 530
```

```
                     ||||||||||||||||||| ||| || :|||| |  |:|:|||||| ||:|
Db        477 CKNVVPVYDLLLEMLNAHTLRGNKSLVTPSPPSPMEESESKEGSQEPQAQ 529


RESULT 8
ESR2_STUVU
ID   ESR2_STUVU     STANDARD;      PRT;   554 AA.
AC   Q9PVE2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta)
GN   ESR2 OR NR3A2.
OS   Sturnus vulgaris (Starling).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata, Euteleostomi;
OC   Archosauria; Aves; Neognathae, Passeriformes, Sturnidae; Sturnus.
OX   NCBI_TaxID=9172;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99427866; PubMed=10499520;
RA   Bernard D.J., Bentley G.E., Balthazart J., Turek F.W., Ball G.F.;
RT   "Androgen receptor, estrogen receptor alpha, and estrogen receptor
RT   beta show distinct patterns of expression in forebrain song control
RT   nuclei of European starlings.";
RL   Endocrinology 140:4633-4643(1999).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF ER-
CC       ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC       LOCALLY SYNTHESIZED ESTROGENS MAY ACT VIA ER BETA, IN ADDITION TO
CC       ER ALPHA, TO MEDIATE SEASONAL OR DEVELOPMENTAL EFFECTS ON NEARBY
CC       SONG NUCLEI.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: BRAIN, PITUITARY, SKELETAL MUSCLE, LIVER,
CC       ADRENAL, KIDNEY, INTESTINE, AND OVARY.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS. A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF113513; AAD56593.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN       25    173       MODULATING.
FT   DNA_BIND     174   239       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      174   194       C4-TYPE.
FT   ZN_FING      210   234       C4 TYPE.
FT   DOMAIN       240   554       STEROID-BINDING.
SQ   SEQUENCE   554 AA;  62173 MW;  EC13A961CEBE59C9 CRC64;

  Query Match             74.9%;  Score 2101;  DB 1;  Length 554;
  Best Local Similarity   74.8%;  Pred. No. 2.1e-161;
  Matches  397;  Conservative   50;  Mismatches   80;  Indels    4;  Gaps    3;

Qy          2 DIKNSPSSLNSPSSYNQSQSTLPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSN 61
              :|||||: : ||: |:|:|| |  ||  :|||||||::| ||| | | | |||::||:| ||
Db         26 EIKNSPAGVISPAPYSQNQSILIAEHSPVYIPSSYMESRHEYSTMAFCSPAMVNYNIASN 85

Qy         62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNR 121
                 : |    |||:|| ||  ||||||| :| | | |||| || ||||| :|  || :|
Db         86 FGDPEAVAARQTSSPGALWSAPGHLSPLSLHCQSSLLYAEQPKSLWCEARPMEPVLPGSR 145

Qy        122 FTLKRKVSGNRCASPV-TGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||| :|| | ||:   ||||:|||||||||||||||||||||||||||||||||||||
Db        146 FTLKRKINGNDCTSPIANNPGSKKDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 205

Qy        181 NDYICPAINQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              |||||||||||||||||||||||||||||||||:|||||||||||:|  | |:|::|
Db        206 NDYICPAINQCTIDKNRRKSCQACRLRKCYEVGMMKCGSRRERCGYRLLRSHRGAEERVH 265

Qy        241 CAGKAKRSGGHAPRVKELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              | :|.|    | ||:|:||  :|||| ||||||||||||:|||| ||||||||||||
Db        266 CLGPARPYSFAATRVKFILLSTVSPFQFVLTLLEAPPPHVLVSRPSKPFTEASMMMSLTK 325

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              |||||||||| |||||||::|||:|||||||||||||:|||||||||||||||||||||
Db        326 LADKELVHMICWAKKIPGFIDLSLYDQVRLLESCWMEVLMIGLMWRSIDHPGKLIFAPDL 385

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              |||||||||||||||||||||||| ||||||||||||||||||||||||||:||  : ::
Db        386 VLDRDEGKCVEGILEIFDMLLAMTSKFRELKLQHKEYLCVKAMILLNSSMFPL--SAEEP 443

Qy        421 DSSPKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
              :|:||| |||| ||:|||||||||| |||| |||||||||||||||||||||||:||
Db        444 ESNPKIHHLLNVVTEALVWVIAKSGIPSQQQTTRLANLLMLLSHVRHASNKGMEHLLSMK 503

Qy        481 CKNVVPVYDLILEMLNAHVLRGCKSSITGSEDSPAEDSKSEEGSQNPQSQ 530
              |||||||||||||||||| ||| :  : |  |  | |  :  |  :  : |
Db        504 CKNVVPVYDLILEMLNAHTLRGQEPLATPEEEGSILEQMEEASSLEKGPPQ 554


RESULT 9
ESR2_COTJA
ID   ESR2_COTJA     STANDARD;      PRT;   472 AA.
AC   O93511;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Coturnix coturnix japonica (Japanese quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=93934;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Foidart A., Lakaye B., Grisar T., Ball G.F., Balthazart J.;
RT   "Sequence and neuroanatomical distribution of estrogen receptor beta
RT   in the quail brain.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 141-286 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=98430913; PubMed=9760113;
RA   Lakaye B., Foidart A., Grisar T., Balthazart J.;
RT   "Partial cloning and distribution of estrogen receptor beta in the
RT   avian brain.";
RL   NeuroReport 9:2743-2748(1998).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
```

```
CC     !- SUBCELLULAR LOCATION: Nuclear.
CC     !  TISSUE SPECIFICITY: A HIGH EXPRESSION IS SEEN IN THE
CC        TELENCEPHALON, DIENCEPHALON, PITUITARY, TESTIS AND KIDNEYS BUT
CC        LITTLE OR NO EXPRESSION IS SEEN IN THE CEREBELLUM, PECTORAL MUSCLE
CC        AND ADRENAL GLAND.
CC     !  DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC        A DNA BINDING DOMAIN AND A C-TERMINAL STEROID BINDING DOMAIN.
CC     !- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC        NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------- --
DR   EMBL; AF045149; AAC36463.2; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid binding.
FT   DOMAIN        1    104       MODULATING.
FT   DNA_BIND    105    170       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     105    125       C4-TYPE.
FT   ZN_FING     141    165       C4-TYPE.
FT   DOMAIN      171    472       STEROID-BINDING.
SQ   SEQUENCE   472 AA;  53411 MW;  DF7A78F0FDBD18BD CRC64;

  Query Match             69.8%;  Score 1959;  DB 1;  Length 472;
  Best Local Similarity   80.1%;  Pred. No. 4.5e-150;
  Matches  370;  Conservative   39;  Mismatches   47;  Indels    6;  Gaps    3;

Qy         46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
              | | |||:|||:| ||  : :    |||:||::|| |||||||  :| ||| |||| ||
Db          1 MAPCSPAMMNYNIASNPGDSESASVRQTSSPSLLWSAPGHLSPLTLHCQLSLLYAEQPKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPV-TGPGSKRDAHFCAVCSDYASGYHYGVW 164
              |||||| ||  |||:|||||||| :|: | ||: : |||||||||||||||||||||||
Db         61 PWCEARPLEPVLPVSRETLKRKTNGSDCTSPIASNPGSKRDAHFCAVCSDYASGYHYGVW 120

Qy        165 SCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRC 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        121 SCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRC 180

Qy        225 GYRLVRRQRSADFQLHCAGKAFRSGGHAPPVPFTTIDASPFQIVTTLLPAFPPHVISP 284
              |||:|| |:::: | || |:  | ||:|:||  :|||| ||||||||||:||:||
Db        181 GYRILKRHRNSFD-- CMGKTKRYNEAAIRVKRTLLSIVSPEQPVLILLEAEPPNVLYSE 237

Qy        285 PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLM 344
              || |||||||||||||||||||| ||||||::|||:|||||||||||||||||||||||
Db        238 PSKPFTEASMMMSLTKLADKELVHMICWAKKIPGFIDLSLYDQVRLLESCWLEVLMIGLM 297

Qy        345 WRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMI 404
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        298 WRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLAMTSRFRELKLQHKEYLCVKAMI 357

Qy        405 LLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQADRLANLLMILSH 464
              |||||||:|| : :: :|:|||| |||| ||||||||||||| |||| ||||||||||
Db        358 LLNSSMEPL--SAEEPESNRKLIHILLNVVTDALVWVIAKSGISPQQQTDRLANLLMILSH 415

Qy        465 VRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSS 506
              |||||||||||:|||||||||||||||||| ||| : |
Db        416 VRHASNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGQRKS 457
```

```
RESULT 10
ESR2_CHICK
ID   ESR2_CHICK     STANDARD;      PRT;   472 AA.
AC   Q98TU5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta) (cERb).
GN   ESR2 OR NR3A2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN White leghorn; TISSUE Testis;
RA   Suzuki M., Mizuno S., Nakabayashi O.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC    !  DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------- --
DR   EMBL; AB036415; BAA88667.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    104       MODULATING.
FT   DNA_BIND    105    170       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING     105    125       C4 TYPE.
FT   ZN_FING     141    165       C4 TYPE.
FT   DOMAIN      171    472       STEROID-BINDING.
SQ   SEQUENCE   472 AA;  53439 MW;  A8CBB8B8CAC65650 CRC64;

  Query Match             69.5%;  Score 1948.5;  DB 1;  Length 472;
  Best Local Similarity   79.3%;  Pred. No. 8.1e-149;
  Matches  371;  Conservative   38;  Mismatches   52;  Indels    7;  Gaps    4;

Qy         46 MTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKS 105
              | | |||:|||:| ||  : :    |||:||::|| |||||||  :| | | |||| ||
Db          1 MAPCSPAMLNYNIASNPGDSESASVRQTSSPSVLWSAPGHLSPLTLHCQSSLLYAEQPKS 60

Qy        106 PWCEARSLEHTLPVNRETLKRKVSGNRCASPV-TGPGSKRDAHFCAVCSDYASGYHYGVW 164
              |||| | ||  ||| ||||||| :|: | ||: : ||||||||||||||||||||||||
Db         61 PWCEVRPLDPVLPVTRETLKRKTNGSDCTSPIASNPGSKRDAHFCAVCSDYASGYHYGVW 120
```

```
Qy        165 SCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSA 234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSA 190

Qy        225 GYPLVPPQPSADFQLHCAGKAKPSGCHAPPVPFLIIDAISPPQLVTLLEAEPPHVLISR 284
              |||::|| |::::   | || |:   | ||:|:||  :|||| ||||||||||:||:||
Db        191 GYRILRPHRNSFD   CMGKTKKYNEAATPVKFLFIISIVSPEQFVTLLEAEPPNVLVSR 247

Qy        285 PSAPFTEASMMMSLIKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLM 344
              || ||||||||||||||||||||||| |||||||::|||:|||||||||||||||||:|||
Db        248 PSKPFTEASMMMSLTKLADKELVHMIGWAKKIPGFIDLSLYDQVRLLESCWMEVLMIGLM 307

Qy        345 WRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMI 404
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        308 WRSIDHPGKLIFAPDLVLDRDPGKCVEGILEIFDMLLAMTSRFRELKLQHKEYLCVKAMI 367

Qy        405 LLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSH 464
              ||||||:||  : :: :|:||| |||| ||||||||||||||| |||| ||||||||||
Db        368 LLNSSMFPL--SPEFPESNRKIHHLLNVVTDALVWVIAKSGIPSQQQTIRLANLLMLLSH 425

Qy        465 VRHASNKGMEHLLSMKCKNVVPVYDLLLEMLNAHVLRG-CKSSITGSE 511
              ||||||||||||:|||||||||||||||||||||||| ||| || :|  |
Db        426 VRHASNKGMEHLLSMKCKNVVPVYDLLLEMLNAHTLRGQRKSPVTHPE 473


RESULT 11
ESR2_ANGJA
ID   ESR2_ANGJA     STANDARD;      PRT;   573 AA.
AC   O13012;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Anguilla japonica (Japanese eel).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC   Anguilla.
OX   NCBI_TaxID=7937;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=96386022; PubMed=8793852;
RA   Todo T., Adachi S., Yamauchi K.;
RT   "Molecular cloning and characterization of Japanese eel estrogen
RT   receptor cDNA.";
RL   Mol. Cell. Endocrinol. 119:37-45(1996).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: LIVER
CC   -!- INDUCTION: BY 17-BETA-ESTRADIOL.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS. A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------- --
DR   EMBL; AB003456; BAA19851.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN       15    170       MODULATING.
FT   DNA_BIND     171   236       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      171   191       C4-TYPE.
FT   ZN_FING      207   231       C4-TYPE.
FT   DOMAIN       237   573       STEROID-BINDING.
SQ   SEQUENCE   573 AA;  63420 MW;  9C64C1D8D39ED40C CRC64;

  Query Match             55.2%;  Score 1549;  DB 1;  Length 573;
  Best Local Similarity   57.7%;  Pred. No. 5.6e-117;
  Matches  313;  Conservative   62;  Mismatches  121;  Indels   46;  Gaps   10;

Qy         13 PSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVINLEGGP--- 69
              |: || :   | :|  :: ||| | || ||:| |:|||||| ::::          |||
Db         83 PTMYNGALPALSMESHAVCIPSPYTDSSHDYAALTFYSPPILSH----------GGPAVP 82

Qy         70 ----GRQTTSPNVLWPT---PGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHILPVNKE 122
                  ||: ||:: ||     ||:||| :| |   :| ||  ||| | : |||      :
Db         83 ESPAARQSLSPSLFWPAHGHHGHVSPLALHFQQPLVYREPAHSPWAEPKPLEH-----GQ 137

Qy        123 TLKPKVSCNPCASPVTGPGS-------KPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKR 175
                |::| | |    | |        | | |||||| ||||||||||||||||||||||
Db        138 AQISKLAGKPMAESPEGISSVGGCPAGKGDMHFCAVCHDYASGYHYGVWSCEGCKAFFKR 197

Qy        176 SIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSA 235
              |||||| ||||||||||||||||||||||||||||||||||:||| ||||| || | :|
Db        198 SIQGHNGYICPATNQCTIDKNRRKSCQACRLRKCYEVGMMKCGVRRERCTYRGAHRRMP 257

Qy        236 D--EQLHCAGKA---KRSCGHAPRVRELLLDALSPEQLVLTLLEAEPPHV-LISRPSAPF 289
              |    | |   :|  |  |: :|    ||:|||||  ::|||||  : |:     ||
Db        258 HIRELAGTGGGARTQRRGPGVVPQTQEAQSSALTPPQLINRTIEAEPPPFIYIMKPLKKPF 317

Qy        290 TEASMMMSLIKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSID 349
              || ||||||| ||||||| ||||||||||||| | ||| ||| ||:|||||||||:|
Db        318 TEDSMMMSLINLADKELVIMISWAKKIPGFVELDLSDQVHLLECCWLEVLMLGLMWRSVD 377

Qy        350 HPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSS 409
              |||||||:||| |:|||| |||||||||||:|| |||||||||| :||:|:||:|||| :
Db        378 HPGKLIFSPDLKLNRDEGSCVEGILEIFDMVLAATSRFRELKLQREEYVCLKAIILLNPN 437

Qy        410 MYPLVTAT-QDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA 468
              :    :   :: :|  || |:|:::||||||| ||| |:: |||| |||:|||||:|:||
Db        438 LCTTSSENPEELESRNKLLHMLDSVIDALVWIIAKKGLTTQQQSARLAHILMILAHIRHL 497

Qy        469 SNKGMEHLLNMKCKNVVPVYDLLIEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQ 528
              |||||||| ||| ||||:|||||||:| :         | | :  |:   || |
Db        498 SNKGMEHLSNMKSKNVVPLYDLLLEMLDANIMH-------SSRMSASYSSQFSPWSQAAQ 550

Qy        529 SQ 530
              ||
Db        551 SQ 552


RESULT 12
ERB1_CARAU
ID   ERB1_CARAU     STANDARD;      PRT;   568 AA.
AC   Q9W669;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Estrogen receptor beta-1 (ER-beta-1).
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Actinopteryqii; Neopteryqii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Carassius.
OX   NCBI_TaxID=7957;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=99169117; PubMed=10068500;
RA   Tchoudakova A.V., Pathak S., Callard G.V.;
RT   "Molecular cloninq of an estroqen receptor beta subtype from the
RT   qoldfish, Carassius auratus.";
RL   Gen. Comp. Endocrinol. 114:388-400(1999).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC    !  SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!  SUBCELLULAR LOCATION: Nuclear.
CC    !  DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC    !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- -- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- -- ---------- ---------- ---------- ---------- ----------
DR   EMBL; AF061269; AAD26921.1;  -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription requlation; DNA-bindinq; Nuclear protein;
KW   Zinc finqer; Steroid-bindinq.
FT   DOMAIN        12    169       MODULATING.
FT   DNA_BIND     170    235       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING      170    190       C4-TYPE.
FT   ZN_FING      206    230       C4-TYPE.
FT   DOMAIN       236    568       STEROID-BINDING.
SQ   SEQUENCE   568 AA;  63539 MW;  2D678D0775DB76F9 CRC64;

  Query Match          52.4%;  Score 1466;  DB 1;  Length 568;
  Best Local Similarity   56.0%;  Pred. No. 2.7e-110;
  Matches  290;  Conservative   75;  Mismatches  127;  Indels   26;  Gaps   10;

Qy         12 SPSSYNCSQSILPLEHCSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPG- 70
              || ::| |   ||:|   | ||| | |  |::   : |||::: |       : |   |
Db         48 SP-TFNSSSPSLPVESHPICIPSPYTDLGHDFTTLPFYSPSLLGY---GTSPLSDCPSV 92

Qy         71 RQTTSPNVIWPTPGHLSDLVVHRQLSHLYA EPQKSPWCEAPSLEHTLPVNPETLKRKVS 129
              || : ||  : ||   |:| | :|:|: : |    |    | |    :|    | : | ::|:
Db         93 RQSLSPTLFWPPHSHVSSLALHQQQTRLQPNHPTGGTWAELTPHDHGEEENCKPLSKRVA 152

Qy        130 GNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATN 189
                 |    | | |:||||||||||||||||||||||||||||||||||||||||||||||
Db        153 VAEETS-TSLRGKADMHYCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATN 210

Qy        190 QCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQ-LHCACKAKR 247
              ||||||||||||||||||||||||||:||| ||:|  |    :|| | :      :|: :
Db        211 QCTIDKNRRKSCQACRLRKCYEVGMMKCGLRRDRSSY----QQRCAQQNRELTRFSGRMRT 266

Qy        248 SGGHAPRVR-  ---    ELLLDALSPEQLVLTLLEAEPPHV-LISRPSAPFTEASMMMS 297
              ||  :  ::          :::  |||||:|:  :::|||| : |:     |||||::|||
Db        267 SGPRSQEIKTVQRPLSGNKVVTMALSPEELIARIMDAEPPEIYLMKDVKKPFTEANVMMS 326

Qy        298 LTKLADKELVHMISWAKKIPGFVELSLEDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFA 357
              || |||||||||||||||||||||: ||||| ||| ||:|||:|||||::|||||:|:
Db        327 LTNLADKELVHMISWAKKIPGFVELGLEDQVHLLECCWLEVLMLGLMWRSVNHPGKLVFS 386

Qy        358 PDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSM YPLVTA 416
              ||| | |||| ||:|  |||||||| ||||||||||| :|| |:|||||||:|
Db        387 PDLSLSRDEGSCVQGFAEIFDMLLAATSRFRELKLQREEYACLKAMILLNSNMCLSSAEG 446

Qy        417 TQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHL 476
              ::  |  ||  ||::|||||| |:|:|:| ||:| |||:|||||||:|| ||||:|
Db        447 GEELQSRSKLLCLLDSVTDALVWAISKTGLSFQQRSTRLAHLLMLLSHIRHVSNKGMDHL 506

Qy        477 LNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSP 514
              :|| | :||:|||||||||:||:: | : | :|  :  |
Db        507 HSMKMKKMVPLYDLLLEMLDAHIMHGSRLSHSGPQADP 544


RESULT 13
ESR2_ONCMY
ID   ESR2_ONCMY     STANDARD;      PRT;   568 AA.
AC   P57782;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estroqen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo qairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopteryqii; Neopteryqii; Teleostei; Euteleostei;
OC   Protacanthopteryqii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hauqq M., Ackermann G., Font K.;
RT   "Molecular cloninq of an estroqen receptor beta subtype from rainbow
RT   trout.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC    !- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC    !- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- -- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- -- ---------- ---------- ---------- ---------- ----------
DR   EMBL; AJ289883; CAC06714.1;  -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription requlation; DNA-bindinq; Nuclear protein;
KW   Zinc-finqer; Steroid-bindinq.
FT   DOMAIN        1    177       MODULATING.
```

```
FT   DNA_BIND     178     243       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      178     198       C4-TYPE.
FT   ZN_FING      214     238       C4-TYPE.
FT   DOMAIN       244     568       STEROID-BINDING.
SQ   SEQUENCE   568 AA;  63813 MW;  F7A7BD7B3B2C2804 CRC64;

  Query Match             51.4%;  Score 1440.5;  DB 1;  Length 568;
  Best Local Similarity   54.9%;  Pred. No. 3e-108;
  Matches  289;  Conservative   85;  Mismatches  121;  Indels   31;  Gaps   11;

Qy         16 YNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNY-SIPSNVTNLEGGPGRQTT 74
              :| |   ||:|   : | | | |  |::  : ||||:: | : | :    |    ||:
Db         49 FNSSSPSLPVESHPVCIQSPYTDLGHDFTTLPFYSPALLGYGTSPLS----ECSSVRQSL 104

Qy         75 SPNVLWPTPGHLSPLVVHPQISHLYA-EPQKSPWCEAPSIFHTLPVNPFTLKPEVSCNPC 133
              || : ||    :| | :|:| : |    |    | |    :|:    |: | :::
Db        105 SPTLPWPPHSQVSSIALHQQHTFIQQNHPTCGTWTELTPHDHSIFFYEPECVFEVAGAIE 164

Qy        134 ASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTI 193
              | |    | | |:||||||||||||||||||||||||||||||||||||||||||||
Db        165 TS--TSLRGKADMHYCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTI 222

Qy        194 DKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRMLKHKRQRDSAEEG--LHCAGKAKESGCH 251
              ||||||||||||||||||||:||| ||:|  |    :|| | ::     :|: : ||
Db        223 DKNRRKSCQACRLRKCYEVGMMKGCLRKDRRGSY    QQRGAQQEELAEPSCPMETSGPE 278

Qy        252 APRVR---------ELLLDALSPEQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLTKL 301
              :  ::           |::  |||:|:|:  :::|||| : |:     |||||:|||| |
Db        279 SQEMKSVPGPLSGNPVVVNMAITPPFLIAPIMDAEPPEIYILMKDMKKPFTEANVMMSLTNL 338

Qy        302 ADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLV 361
              |||||||||||||:|||||||||||| ||| ||:|||:|||||:|||||||||||||| |
Db        339 ADKELVHMISWAKKVPGFVDLSLFDQVHLLECCWLEVLMLGLMWRSVNHPGKLIFSPDLS 398

Qy        362 LDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDAD 421
              | |||| ||:| :|||||||| |||||||||| :||:|:|||||||:|   ::::: :
Db        399 LSRDEGSCVQGFVEIFDMLLAATSRFRELKLQHFEYVCLKAMILLNSNM--CLSSSEQCE 456

Qy        422 SSR---KLAHLLNAVTDALVWVIAKSGISSQQQSMELANLLMLLSHVRHASNKGMEHLLN 478
              :   ||  ||:|||||||| |:|:|:| ||:| |||:|||||||:|| ||||||||
Db        457 ELQRRSKLLCLLDSVTDALVWAISKTGLSPQQRSTRLAHLLMLLSHIRHLSNKGMDHLHC 516

Qy        479 MKCKNVVPVYDLLLEMLDAHVLRGCKSSITGSECSPAEDSKSKEGS 524
              || | :||:|||||||||:||:: | : | :|   :| | :  :| :
Db        517 MKMKKMVPLYDLLLEMLDAHIMHGSRLSHSGP--APKESTGVQEAT 560


RESULT 14
ESR2_MICUN
ID   ESR2_MICUN     STANDARD;      PRT;   673 AA.
AC   P57781;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Micropogonias undulatus (Atlantic croaker).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
OC   Sciaenidae; Micropogonias.
OX   NCBI_TaxID=29154;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=20461442; PubMed=11005855;
RA   Hawkins M.B., Thornton J.W., Crews D., Skipper J.K., Dotte A.,
RA   Thomas P.;
RT   "Identification of a third distinct estrogen receptor and
RT   reclassification of estrogen receptors in teleosts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:10751-10756(2000).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: ABUNDANT IN THE LIVER AND TESTES, LESS
CC       ABUNDANT IN THE OVARY AND BARELY DETECTABLE IN THE MUSCLE.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF298181; AAG16711.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN         1     181       MODULATING.
FT   DNA_BIND     182     247       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING      182     202       C4-TYPE.
FT   ZN_FING      218     242       C4-TYPE.
FT   DOMAIN       248     653       STEROID-BINDING.
SQ   SEQUENCE   673 AA;  74680 MW;  28394517577E3D01 CRC64;

  Query Match             51.0%;  Score 1431;  DB 1;  Length 673;
  Best Local Similarity   59.5%;  Pred. No. 2.2e-107;
  Matches  303;  Conservative   76;  Mismatches  125;  Indels   62;  Gaps   18;

Qy          6 SPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEY---PAMTFYSPAVMNYSIPSN 61
              ||  |  |: |:     | ::  :: ||| | || |||       :|||||:::|| |
Db         31 SPGLL--PAVYS---PPLGMDSHTVCIPSPYIDSSHEYNHSHGPLTFYSPSVLSYSRPP- 84

Qy         62 VTNLPSGPGF-QTTSPNVLWPTPGH-ISPLVVHPQISHLYAEPQ-KSPWCEAK----- 111
              :||    |     : ||:  ||:  |  :  | :|   | :| ||   :|| |::
Db         85 ITN---SPSSLCPSLSPSAFWPSHNHPTMPSLLLHCPESIVYNEPSPHAPWLESKAHSIN 141

Qy        112 -SLEHTLPVNPETLKPEVSC------NRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVW 164
               |    :  |:  :||  |      : |:| |     | | |||||| ||||||||||
Db        142 ASSSSTLCCNKSLVKPSEFGVEDMNSSLCSSAV----GKADMHFCAVCHDYASGYHYGVW 197

Qy        165 SCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRR 224
              |||||||||||||||||||||||||||||||||||||||||||||||||:||| |||||
Db        198 SCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGVFKDRR 257

Qy        225 GYRLVRRQRS-------ADEQLHCAGKAKRSGCH-- ------ -AP -RVRELLLDALSP 265
              ||  | :|             |   |   |: |          ||  : : : | |:||
Db        258 SYRGARHRRGGLQPRDPTGRGLVRVGLGSPAQRHLHLEAPLTPLAPILQAKRVRLSAMSP 317

Qy        266 EQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSL 324
              |: :  :::|||| : |:     |||||||||||| |||||| ||||||||||||||||
Db        318 EEFISRIMDAEPPEIYLMKDLKKPFTEASMMMSLTNLADKELVHMISWAKKIPGFVELSL 377

Qy        325 FDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATT 384
              || || ||:|:||:|||||:|||||||:||  |:|:||:|||||:|||||||| |
Db        378 ADQINLLKCCWLEILMIGLMWRSVDHPGKLIFSPDLKLNREEGQCVEGIMEIFDMLLAGT 437
```

```
Qy          385 SREFRELKLQHKEYLCVKAMILLNSSMYPLVTAT QDADSSPELAHLLNAVTDALVWVIAK 444
                ||||||||||| ||||||||||||||||::       | :: :|  ||  ||:|| ||||| ||| 
Db          438 SREFRELKLQREEYVCLKAMILLNSNLCTSSPQTAEELESRNKLLRLLDSVIDALVWAISK 497

Qy          445 SGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGC 504
                |:::|||:::|| :| ||||||:|| ||||||:||  || |||| ||||||||||:|:   |
Db          498 MGLTTQQQTLRLGHLTMLLSHIRHVSNKGMDHLSTMKRKNVVLVYDLLLEMLDANTSSG- 556

Qy          505 KSSITGSECSPAEDSKS KEGSQNPQ 528
                ||: | :  |:: :  | ||
Db          557 -GSQPSSSPSSETYSDQHQYPQ 577


RESULT 15
ESR2_MACMU
ID   ESR2_MACMU     STANDARD;      PRT;   279 AA.
AC   Q9TTE5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER beta) (Fragment).
GN   ESR2 OR NR3A2.
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9544;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20113291; PubMed=10644527;
RA   Wu W.X., Ma X.H., Smith G.C.S., Nathanielsz P.W.;
RT   "Differential distribution of ERalpha and ERbeta mRNA in intrauterine
RT   tissues of the pregnant rhesus monkey.";
RL   Am. J. Physiol. 278:C190-C198(2000).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- ---------- ---------- ---------- ---------- ---------- ---------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ---------- ---------
DR   EMBL; AF119229; AAD54069.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   SMART; SM00430; HOLI; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; PARTIAL.
KW   Receptor; Transcription regulation; DNA binding; Nuclear protein;
KW   Zinc finger; Steroid binding; Phosphorylation.
FT   NON_TER       1       1
FT   DOMAIN        -1     >279       STEROID-BINDING.
FT   MOD_RES      251     251       PHOSPHORYLATION (BY SIMILARITY).
FT   NON_TER      279     279
SQ   SEQUENCE   279 AA;  31105 MW;  858D9B7D01DA0401 CRC64;

  Query Match             49.8%;  Score 1398;  DB 1;  Length 279;
  Best Local Similarity   98.2%;  Pred. No. 4.1e-105;
  Matches  274;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          238 QLRCAGKAKRSGCHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMS 297
                ||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db            1 QLRCAGKAKRSGSHAPLVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMS 60

Qy          298 LTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFA 357
                ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db           61 LTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMVGLMWKSIDHPGKLIFA 120

Qy          358 PDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSNMYPLVTAT 417
                ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db          121 PDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSNMYPLVTAT 180

Qy          418 QDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLL 477
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 QDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLL 240

Qy          478 NMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAE 516
                :||||||||||||||||||||||||||||||||||||||
Db          241 SMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAE 279


RESULT 16
ESR2_SPAAU
ID   ESR2_SPAAU     STANDARD;      PRT;   559 AA.
AC   Q9W6M2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Sparus aurata (Gilthead sea bream).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
OC   Sparidae; Sparus.
OX   NCBI_TaxID=8175;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Socorro S., Power D.M., Olsson P.-E., Canario A.V.M.;
RT   "Two estrogen receptors expressed in teleost fish: cDNA cloning,
RT   characterization and tissue distribution.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- ---------- ---------- ---------- ---------- ---------- ---------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ---------- ---------
DR   EMBL; AF136980; AAD31033.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
```

```
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid binding.
FT   DOMAIN          1     155       MODULATING.
FT   DNA_BIND      156     221       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING       156     176       C4-TYPE.
FT   ZN_FING       192     216       C4-TYPE.
FT   DOMAIN        222     559       STEROID BINDING.
SQ   SEQUENCE   559 AA;  62698 MW;  656123B4B721FF76 CRC64;

  Query Match             49.4%;  Score 1385.5;  DB 1;  Length 559;
  Best Local Similarity   54.0%;  Pred. No. 7.8e-104;
  Matches  276;  Conservative   67;  Mismatches  145;  Indels   33;  Gaps    8;

Qy          3 IKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNV 62
              : :||   | |                 | ||| | |  |::| : |||    :|: |  :
Db         29 VLSSPMETNQP----------------ICIPSPYIDRGHDFPTIPFYSAINFSYANPPAI 72

Qy         63 TNLEGGPGRQTTSPNVIWPIPGHISPIVVHRQHSHLYAFPQ--  KSFWCTAE-SPHTI 117
              ::      || ||:: ||: ||:   :    | || | ||     :||| |   |:: |
Db         73 SDRPS- VHQILSPSLFWPSHGHVGTTL---PLHHLQARPQHGQAVQSPWVELSPLDNVL 127

Qy        118 PVNRFTIKRKVSGNPCASPVTGPGSKPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSI 177
              ::   :|              | | | |||||| ||||||||||||||||||||||||||
Db        128 TSSKSARRPSQENE---EGFVSSGCKADLHFCAVCHDYASGYHYGVWSCEGCKAFFKRSI 184

Qy        178 QGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRM--LVKRQRSA 235
              | ||||||||||||||||||||||||||| ||| ||| ||| |:|| :|   :||
Db        185 QRHNDYICPATNQCTIDKNRRKSCQACRLHKCYNVGMIKGGMRKERRGNFPDPQMRRVTEL 244

Qy        236 DPQLHCAGKAKPSGGHAPVVPELLLDALSPEQIVLTLLEAEPPHV LISRPSAPFTEASM 294
              |   :| :  :|      :       || :||:  ::|||| : |:     | |||::
Db        245 SSQGRTSGPSVLNGPAVGPLNTPQPPALTSKQLIERIMEAEPPEIYLMKDMRRPLFEANI 304

Qy        295 MMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKL 354
              ||||| |||||||||||:||||||||:|| | ||| ||| ||:||||:|||||||:|||||
Db        305 MMSLTNLADKELVHMITWAKKIPGFLELGLLDQVHLLECCWLEVLMIGLMWRSVDHPGKL 364

Qy        355 IFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSM-YPL 413
              ||:||| | |:|| ||:| |||||||:| ||| |||||| :|||:|:|||||||:|
Db        365 IFSPDLSLSRFEGSCVQGFLEIFDMLLAATSRVFRLSWQEFYVCLKAMILLNSNMCLES 424

Qy        414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 473
              :::  |  ||  ||:|||||||| |||:|:: :|  |||:|||||||:|| |||||
Db        425 SEGSEELQSRSKLLRLLDAVTDALVWAIAKTGLTFQQQYTRLAHLLMLLSHIRHVSNKGM 484

Qy        474 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCK 504
              :||  || ||:||:||||||||:||::  :
Db        485 DHLHGMKMKNMVPLYDLLLEMLDAHRMHSSR 515


RESULT 17
ESR2_ICTPU
ID   ESR2_ICTPU     STANDARD.      PRT,   575 AA.
AC   Q91AK1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2
OS   Ictalurus punctatus (Channel catfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC   Ictaluridae; Ictalurus.
OX   NCBI_TaxID=7998;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=20218982; PubMed=10753576;
RA   Xia X., Gale W.L., Chang X., Langenau D., Patino R., Maule A.G.;
RA   Densmore L.D.;
RT   "Phylogenetic sequence analysis, recombinant expression, and tissue
RT   distribution of a channel catfish estrogen receptor beta.";
RL   Gen. Comp. Endocrinol. 118:139-149(2000).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: OVARY AND TESTIS.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF185568; AAF63157.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN          1     160       MODULATING.
FT   DNA_BIND      161     226       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING       161     181       C4-TYPE.
FT   ZN_FING       197     221       C4-TYPE.
FT   DOMAIN        227     575       STEROID-BINDING.
SQ   SEQUENCE   575 AA;  63907 MW;  E3863C6AB8EB824F CRC64;

  Query Match             49.2%;  Score 1381;  DB 1;  Length 575;
  Best Local Similarity   53.9%;  Pred. No. 1.9e-103;
  Matches  297;  Conservative   64;  Mismatches  112;  Indels   78;  Gaps   15;

Qy         13 PSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSNVINLEGGPGRQ 72
              || |     : | :: |:: ||| | |: |          | |::  ||   |   |
Db         25 PSHY---PTAIGVDRGTVCIPSPYADNGH----------AEMSFCGPSAPENPAIAP--- 68

Qy         73 TISPNVLWPI--PGHLSPLVVHPQLSHLYAEPQ-RSPWCEAPSLEHTLIVNPETIEP EV 128
              ||:: | :  |  : | :|   :|||   : | : :   ||   :  | | |:
Db         69 PLSPSLFWSSHNPGAMPALPLHCPPALPYSEPHTHTAWVDTK- PHTSGKHSSPLSRPKL 126

Qy        129 SGNRCASPVTGPGS----------------KRDAHFCAVCSDYASGYHYGVWSCEGCKA 171
               | |   |  | |                  ||| ||| || ||||||||||||||||
Db        127 EGKR---PEDGDGDEALDPDDPSSSSSGAVVKRDMHFCVVCHDYASGYHYGVWSCEGCKA 183

Qy        172 FFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRM-LV 230
              ||||||||||||||||||||||||||||||||||||||:|:|:|||||||||||  ||
Db        184 FFKRSIQGHNEYICPATNQCTIDKNRRKSCQACRLRKCYEMGMMKCGARRDRCGYRASRR 243

Qy        231 KRQRSADPQLHCAGKAKRSGGHAPRVKELLLDALSPEQL 268
                        ||    || :   :|  |         ||||||
Db        244 IAPMRDGSARPVGVRGQSQRLPHTPLHVSLSIPVSRASAESG--------- PSGLSPEQL 294

Qy        269 VTILEAEPPHVLISRPSEPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQ 327
              |  :|||||  : : :    |:|:::|||||:||||| |||||||||||||||  |
Db        295 VYCILEAEPPQILYSKFDQMEPYIESIVMMSLRLADKELVLMISWAKHIPGFVELSLADQ 354

Qy        328 VRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRF 387
              |:||| ||:||||:|||||||||||||: || |:|||| ||||:|||||||| :|||
```

```
Db        455 VQLLECCWLEVLMIGLMWRSIDHPGKLIFSLDLKLNRDEGNCVEGIMEIFDMLLAGSSRF 514

Qy        388 RELKLQHKEYLCVKAMILLNSSMYPLVTATQ-DADSSRKLAHLLNAVTDALVWVIAKSGL 446
              |||||| :||:|:|||:|||||||    :| : | :|  ||  |||:||||||| |:::|:
Db        515 RELKLQKEEYVCLKALILLNSSMYMTSSACEKDLESRTKLLRLLDAVTDALVWAISRTGL 574

Qy        447 SSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL  500
              |:||||  |||:|||||||:|| |||||||| :|: |||| :|||||||:|::   |
Db        575 STQQQSARLAHLLMLLSHIRHLSNKGMEHLSSMKRKNVVPLYDLLLEMLDANMAQSEHVS 634

Qy        501 SSTIGSECSPA 515
              :|:     :||
Db        635 TSVCTDPVTPA 645


RESULT 18
ERB2_CARAU
ID   ERB2_CARAU     STANDARD;      PRT;   610 AA.
AC   Q91AL9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta 2 (ER-beta-2).
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Carassius.
OX   NCBI_TaxID=7957;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Pituitary;
RX   MEDLINE=20246301; PubMed=10786629;
RA   Ma C.H., Dong K.W., Yu K.L.;
RT   "cDNA cloning and expression of a novel estrogen receptor beta subtype
RT   in goldfish (Carassius auratus).";
RL   Biochim. Biophys. Acta 1490:145-152(2000).
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN PITUITARY,
CC       TELENCEPHALON AND HYPOTHALAMUS AS WELL AS IN THE LIVER.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C-TERMINAL STEROID BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- ----- ---- ---------- ---------- ---------- ---------- ------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ----- ---- ---------- ---------- ---------- ---------- ------
DR   EMBL; AF177465; AAF35170.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN       1    170       MODULATING.
FT   DNA-BIND     171    236       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      171    191       C4-TYPE.
FT   ZN_FING      207    231       C4-TYPE.
FT   DOMAIN       237    588       STEROID-BINDING.
SQ   SEQUENCE   610 AA;  67851 MW;  D29F2CDFCC165C67 CRC64;

  Query Match             44.1%;  Score 1378.5;  DB 1;  Length 610;
  Best Local Similarity   51.4%;  Pred. No. 3.2e-103;
  Matches  382;  Conservative   73;  Mismatches  141;  Indels   82;  Gaps   15;

Qy          5 NSPSSLNSPSSYNCSQSTLPLEHGSTYTPSSYVDSHHEYP - - - - - -  AMTFYSPA   52
              ||| | |  |  |    | | ::  :| ||| ||::  :|           |:| |||
Db         21 NSPNIL-PLLYT-SQIGMDSQTICIPSPYVEACQDYSPPHGGEISHGALILYSPVSS 75

Qy         53 -VMNYSTPSNVTNLEGGPGRQTTSP---NVLWPT-PGHL- - - --SPLVVHRQLSHLYA 100
               |: |: |          | ::  |    : ||  | |          |||     :|
Db         76 PVLGYTHP-------- -PVSESLVPLNSAIEWPPHPTHSTPSLHCPSPLAYRETHAHTTW 126

Qy        101 FPQFSPWCFARS-LERTIPVNPFTLPRFVSCNRCASPVTGPGSKRDAHFCAVCSDYASGY 159
              |  |:   :: | | |      : | ::: |:   :|  |   | | |||||| ||||||
Db        127 EDAKTHINQSSSSVLTHA-----KLLGQQLDGDDGELNPSPGTLGKGDTHFCAVCHDYASGY 181

Qy        160 HYGVWSCEGCKAFFKRSIQGHNDYICPAINQCTIDKNRRKSCQACRLRKCYEVGMVKCGS 219
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||:|||
Db        182 HYGVWSCEGCKAFFKRSIQGHNDYICPAINQCTIDKSRRKSCQACRLRKCYEVGMMKCGV 241

Qy        220 RRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLLDA--              - - -  262
              ||||| ||  | ::|   :   :| |     ||:    |  |
Db        242 RRERCSYRGGRHRRNPPIR-DSSGGATGVRGHSQPHLEFPLSPTHPLFPLGDRAEGCGQN 300

Qy        263 LSPEQLVLQTLLEAEPPHVLISRP SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVE 321
              |||||||  :|||||| : :  |   |:|||||||||| |||||| ||||||||||||
Db        301 LSPEQLVNQTLEAEPPQTYLREPTKKPYIEASMMMSLTNLADKELVHMISWAKKIPGFVE 360

Qy        322 LSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLL 381
              |:| ||| ||| ||:::||:|||||:|||||||:||| |:||| |||||:|||||
Db        361 LTLSDQVHLLECCWLDILMLGLMWRSVDHPGKLIFSPDLKLNRDEGTCVEGIMEIFDMLL 420

Qy        382 ATTSRFRFLKLQHKEYVCLKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALVWV 440
              |||||||||||| :||:|:|||||||    :  | :| :|  |:  ||::|||||||||
Db        421 ATTSRFRFLKLQREEYVCLKAMILLNSSNCSRLPQTPEDVFSRGKVIRLLDSVTDALVWT 480

Qy        441 IAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
              |:::|:|| |||:|||:|||||||:|| ||||:|||  || |||| :|||||||:|:|:
Db        481 ISRTGLSSBQQSIRLAHLLMLLSHIRHLSNKGIEHLSTMKRKNVVILYDLLLEMLDANTS 540

Qy        501 RGCKSSTIGSECS- - - - - - - - - - - - - - - - - PAEDSKSKEGSQNPQSQ 540
              :  :    :: |                              ||    :::   :|||::
Db        541 QSSRMLAAHTKASLRMDTQQTFELHFTSKQQPALKESNQDFRHSPQAE 588


RESULT 19
ESR3_MICUN
ID   ESR3_MICUN     STANDARD;      PRT;   565 AA.
AC   P57783;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor gamma (ER-gamma).
GN   ESR3 OR NR3A3.
OS   Micropogonias undulatus (Atlantic croaker).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
OC   Sciaenidae; Micropogonias.
OX   NCBI_TaxID=29154;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=20461442; PubMed=11005855;
RA   Hawkins M.B., Thornton J.W., Crews D., Skipper J.K., Dotte A.,
RA   Thomas P.;
RT   "Identification of a third distinct estrogen receptor and
```

```
RT   reclassification of estrogen receptors in teleosts ";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:10751-10756(2000).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: ABUNDANT IN THE OVARY AND TESTES, BARELY
CC       DETECTABLE IN THE BRAIN AND MUSCLE AND UNDETECTABLE IN THE LIVER.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF298182; AAG16712.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    168       MODULATING.
FT   DNA_BIND    169    234       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     169    189       C4-TYPE.
FT   ZN_FING     205    229       C4-TYPE.
FT   DOMAIN      235    285       HINGE.
FT   DOMAIN      286    520       STEROID-BINDING.
SQ   SEQUENCE   565 AA;  62903 MW;  D9954D2F740CCA5E CRC64;

  Query Match             48.9%;  Score 1373;  DB 1;  Length 565;
  Best Local Similarity   51.1%;  Pred. No. 8e-103;
  Matches  283;  Conservative   70;  Mismatches  139;  Indels   62;  Gaps   11;

Qy          14 SSYNCSQSILPLEHGSIYIPSSYVDSHHEYTFAMTIYSPAVMNYSIP  SNVTNLEGGPGR 71
               ||   || |          | | | |  |::| : |||| : :|: |  |: |::
Db          35 SSLETSQPIC--------ITSPYTDLGHDFPTIPFYSPTIFSYAGPSISDCTSV-----H 81

Qy          72 QTTSPNVLWPTPCHL-SPLVVHRQLSHLYAEPQKSPWCEARSLEHIL-              117
               |: :|:: || || || :|     : :| :||| |   |-- |
Db          82 QSLNPSLFWPSPCHMCSPTPLHHS---QHCQIPQSFWVFISPLPNVEKIPQDQASLPLAV 138

Qy         118 -PVNRFTLERRVSGNPCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFRRS 176
                ||  :: :|:          |   | | | |:|||| ||||||||||||||||||||||
Db         139 VPVRHKSARRR--SQFSFEAVVTSGGKTDLHYCAVCHDYASGYHYGVWSCEGCKAFFERS 196

Qy         177 IQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYR--LVRHQRS 234
               ||  |:|||||||:|||||||||||||||||||||||| ||| |:|| ||   :|
Db         197 IQRDNEYICPATNECTIDKNRRKSCQACRLRKCYEVGMIKCGMRKERCGNYRSPQMRRMIR 256

Qy         235 ADEQLHCAGKAKRSGGHAPRVRFLLIDAISPFQLVITLIEAEPPHV-LISRPSAPFTEAS 293
               |     :  :|      :       || ||:  |:|||| : |:     | |||
Db         257 LTSQGRTDSSSVETGSAVVSLNAPQPSAITSEQLIERLMEAEPPEIYLMKDMKKPLTEAK 316

Qy         294 MMMSLTKLADKELVHMISWAKKIPGFVELSLEDQVRLLESCWMEVLMMGLMWRSIDHPGK 353
               :|||| |||||||||||:|||||||||| | ||| ||| ||:|||:|||||:|||||
Db         317 VMMSLTNLADKELVHMITWAKKIPGFVELGLLDQVHLLECCWLEVLMVGLMWRSVDHPGK 376

Qy         354 LIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPL 413
               |:|:||| | |:|| ||:|  |||||||| ||| |||||| :||:|:|||||||:|
Db         377 LVFSPDLSLSREEGSCVQGFAEIFDMLLAATSRVRELKLQREEYVCLKAMILLNSNM--- 433

Qy         414 VTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 473
                     ::|| ||  ||:|||||||  | |: :| :||  |||:|||||||:|| |||||
Db         434 --CLSSSESSSKLLRLLDAVTDALVSAIGKTVLSFRQQYTRLAHLLMLLSHIRHVSNKGM 491

Qy         474 EHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDS-------------- 518
               :||  || ||:||:|||||||:||::   :       | | : :
Db         492 DHLHCMKMKNMVPLYDLLLEMLDAHIMHSSRLPPPSPEQEPEDQADAPAPPHSSGSGPSY 551

Qy         519 ----KSKEGSQNPQ 528
                   | ||:  ||
Db         552 TWTPSSSEGAGEPQ 565


RESULT 20
ESR2_ORENI
ID   ESR2_ORENI     STANDARD;      PRT:   557 AA.
AC   Q9YH32;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor beta (ER-beta).
GN   ESR2 OR NR3A2.
OS   Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidei;
OC   Cichlidae; Oreochromis.
OX   NCBI_TaxID=8128;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chang X.T., Kobayashi T., Todo T., Yoshiura Y., Ikeuchi T.,
RA   Kajiura H., Nakamura M., Nagahama Y.;
RT   "cDNA sequence of tilapia type beta estrogen receptor.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS ESTROGENS WITH AN AFFINITY SIMILAR TO THAT OF
CC       ER-ALPHA, AND ACTIVATES EXPRESSION OF REPORTER GENES CONTAINING
CC       ESTROGEN RESPONSE ELEMENTS (ERE) IN AN ESTROGEN-DEPENDENT MANNER.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       ALPHA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U75605; AAB00246.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    154       MODULATING.
```

```
FT   DNA_BIND     155     220        NUCLEAR RECEPTOR TYPE.
FT   ZN_FING      155     175        C4 TYPE.
FT   ZN_FING      191     215        C4 TYPE.
FT   DOMAIN       221     557        STEROID BINDING.
SQ   SEQUENCE   557 AA;  61779 MW;  82A474988C3CCD76 CRC64;

  Query Match             48.2%;  Score 1352;  DB 1;  Length 557;
  Best Local Similarity   52.0%;  Pred. No. 3.9e-101;
  Matches  283;  Conservative   70;  Mismatches  145;  Indels   46;  Gaps   13;

Qy          1 MDIKNSPSSLNSPSSYNCSQSTLDFFHGSTYTPSSYVGSHHFYPAMTFYSFAVMNYSTPS 60
              :|    : | : ||   : |  :       | |   | |  |::  : |||| : :|
Db         17 VDSSRAGSRTLSPTLGSSSPGLSHPTSQPTCIRSPYTTDLGHDFTTIPFYSPTIFSY---- 72

Qy         61 NVTNLEGGPG---RQTTSPNVLWPTPGHL SPLVVH-RQLSHLYAEPQKSPWCEAR 111
                 |||      |: | :: || : | : :|: :|  |      :  ::||
Db         73 ---GGPSISPCSSVHQSISASIFWPSPGRVCTPITLHCPQGPSQQCQSAQTPW-- 122

Qy        112 SLEHTLPVNRETLKRKVSGNPCASPVTGPGSKPDAHFCAVCSDYASGYHYGVWSCEGCKA 171
              :  : ::  :|      :   | | | |:|||| ||||||||||||||||||
Db        123 --DSVITTSKSVRRR--SQESEESMVSSGGKADLHYCAVCHDYASGYHYGVWSCEGCKA 177

Qy        172 FFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSKRERCGYRLVRR 231
              |||||||||||||||||||||||||||||||||||||||||| ||| |:|| ||  :
Db        178 FFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMPKCGTRKERGNYRNSQA 237

Qy        232 QRSADEQLHCAGK AKPSGGHAPRVREL---LLDALSPPQLVETLLEAEPPHV-LISPPS 286
              :|    :|   || |:  |   |     |      ||:|||:  :|||||  : |:
Db        238 RRLT--RLSSQGKIAEPKGTGPGPAPGSLNKPEPKDAITPEQLLEPLLEAEPPEIYVEDAK 295

Qy        287 APFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWR 346
              | ||||:|| || |||||||||||||||||||||||| ||| ||| ||:||||:|||||
Db        296 RPLTEASVMMLLTNLADKELVHMISWAKKIPGFVELSLVDQVHLLECCWLEVLMIGLMWR 355

Qy        347 SIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILL 406
              |:||||||| ||| | |:|| ||:| :|||||:| |:| |||||| :|||:|||||||
Db        356 SVDHPGKLIFCPDLSLSREEGSCVQGFVEIFDMLLAATTRVRELKLQREEYVCLKAMILL 415

Qy        407 NSSM-YPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
              ||:|      :::|  | ||  |||:|||||  | |:|:: :||  |||:|||||||:
Db        416 NSNMCLSSSDCSEDLQSRSKLLRLLDAMTDALVLAIGKTGLTFQQQYTRLAHLLMLLSHI 475

Qy        466 RHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAE--DSKSSKEG 523
              || |||||:||  || ||:||:||||||||:||:          | | | :     ||:
Db        476 RHVSNKGMDHLHCMKMKNIVPLYDLLLEMLDAHTMH------SSCLPHQPPQQDSKIQG 527

Qy        524 SQNP 527
              |: |
Db        528 SEVP 531


RESULT 21
ESR1_HUMAN
ID   ESR1_HUMAN      STANDARD;      PRT;   595 AA.
AC   P03372; Q14276; Q13511; Q9UIS7; Q9UDZ7; Q9N051;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RX   MEDLINE=86122927; PubMed=2753802;
RA   Greene G.L., Gilna P., Waterfield M., Baker A., Hort Y., Shine J.;
RT   "Sequence and expression of human estrogen receptor complementary
RT   DNA.";
RL   Science 231:1150-1154(1986).
RN   [2]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RX   MEDLINE=86146892; PubMed=3754034;
RA   Green S., Walter P., Kumar V., Krust A., Bornert J.-M., Argos P.,
RA   Chambon P.;
RT   "Human oestrogen receptor cDNA: sequence, expression and homology to
RT   v-erb-A.";
RL   Nature 320:134-139(1986).
RN   [4]
RP   SEQUENCE FROM N.A. (LONG ISOFORM), AND VARIANT ASP-411 INS.
RC   TISSUE=Breast;
RX   MEDLINE=96174665; PubMed=8600466;
RA   Pink J.J., Wu S.Q., Wolf D.M., Bilimoria M.M., Jordan V.C.;
RT   "A novel 80 kDa human estrogen receptor containing a duplication of
RT   exons 6 and 7.";
RL   Nucleic Acids Res. 24:962-969(1996).
RN   [4]
RP   SEQUENCE OF 152-595 FROM N.A. (LONG ISOFORM).
RX   MEDLINE=20084372; PubMed=10619354;
RA   Schubert E.L., Lee M.K., Newman B., King M.C.;
RT   "Single nucleotide polymorphisms (SNPs) in the estrogen receptor gene
RT   and breast cancer susceptibility.";
RL   J. Steroid Biochem. Mol. Biol. 71:21-27(1999).
RN   [5]
RP   SEQUENCE OF 216-434 FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=93153765; PubMed=7916651;
RA   Pfeffer U., Fecarotta E., Castagnetta L., Vidali G.;
RT   "Estrogen receptor variant messenger RNA lacking exon 4 in estrogen-
RT   responsive human breast cancer cell lines.";
RL   Cancer Res. 53:741-743(1993).
RN   [6]
RP   SEQUENCE OF 110-117, PHOSPHORYLATION, AND MUTAGENESIS.
RX   MEDLINE=96026869; PubMed=7476978;
RA   Joel P.B., Traish A.M., Lannigan D.A.;
RT   "Estradiol and phorbol ester cause phosphorylation of serine 118 in
RT   the human estrogen receptor.";
RL   Mol. Endocrinol. 9:1041-1052(1995).
RN   [7]
RP   SEQUENCE OF 354-548 FROM N.A.
RC   TISSUE=Breast carcinoma;
RA   Naundorf H., Becker M., Fiebig C., Buettner B., Fichtner I.;
RT   "Mechanisms of acquired tamoxifen resistance in a xenotransplanted
RT   human breast carcinoma.";
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 413-595 FROM N.A.
RA   Parker A.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE OF 532-542, AND PHOSPHORYLATION.
RX   MEDLINE=95280953; PubMed=7539106;
RA   Arnold S.F., Obourn J.D., Jaffe H., Notides A.C.;
RT   "Phosphorylation of the human estrogen receptor on tyrosine 537 in
RT   vivo and by src family tyrosine kinases in vitro.";
RL   Mol. Endocrinol. 9:24-33(1995).
RN   [10]
RP   PHOSPHORYLATION.
RX   MEDLINE=95140025; PubMed=7838153;
RA   Arnold S.F., Obourn J.D., Jaffe H., Notides A.C.;
RT   "Serine 167 is the major estradiol-induced phosphorylation site on the
RT   human estrogen receptor.";
RL   Mol. Endocrinol. 8:1208-1214(1994).
RN   [11]
RP   PHOSPHORYLATION OF SER-104 AND SER-106, AND MUTAGENESIS.
RX   MEDLINE=99357754; PubMed=10428798;
RA   Rogatsky I., Trowbridge J.M., Garabedian M.J.;
RT   "Potentiation of human estrogen receptor alpha transcriptional
RT   activation through phosphorylation of serines 104 and 106 by the
RT   cyclin A-CDK2 complex.";
RL   J. Biol. Chem. 274:22296-22302(1999).
RN   [12]
RP   STRUCTURE BY NMR OF 180-262.
RX   MEDLINE=91061891; PubMed=2247153;
RA   Schwabe J.W.E., Neuhaus D., Rhodes D.;
```

```
RT   "Solution structure of the DNA-binding domain of the oestrogen
RT   receptor.";
RL   Nature 348:458-461(1990).
RN   [13]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 180-262.
RX   MEDLINE=94037104; PubMed=8221895;
RA   Schwabe J.W.E., Chapman L., Finch J.T., Rhodes D.;
RT   "The crystal structure of the estrogen receptor DNA-binding domain
RT   bound to DNA: how receptors discriminate between their response
RT   elements.";
RL   Cell 75:567-578(1993).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 305-548.
RX   MEDLINE=97478539; PubMed=9338790;
RA   Brzozowski A.M., Pike A.C.W., Dauter Z., Hubbard R.E., Bonn T.,
RA   Engstroem O., Oehman L., Greene G.L., Gustafsson J.-A., Carlquist M.;
RT   "Molecular basis of agonism and antagonism in the oestrogen
RT   receptor.";
RL   Nature 389:753-758(1997).
RN   [15]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 306-544.
RX   MEDLINE=98263297; PubMed=9600906;
RA   Tanenbaum D.M., Wang Y., Williams S.P., Sigler P.B.;
RT   "Crystallographic comparison of the estrogen and progesterone
RT   receptor's ligand binding domains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:5998-6003(1998).
RN   [16]
RP   X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 294-554.
RX   MEDLINE=99061051; PubMed=9875847;
RA   Shiau A.K., Barstad D., Loria P.M., Cheng L., Kushner P.J.,
RA   Agard D.A., Greene G.L.;
RT   "The structural basis of estrogen receptor/coactivator recognition
RT   and the antagonism of this interaction by tamoxifen.";
RL   Cell 95:927-937(1998).
RN   [17]
RP   3D-STRUCTURE MODELING OF 311-547.
RX   MEDLINE=98280806; PubMed=9619507;
RA   Maalouf G.J., Xu W., Smith T.J., Mohr S.C.;
RT   "Homology model for the ligand-binding domain of the human estrogen
RT   receptor.";
RL   J. Biomol. Struct. Dyn. 15:841-850(1998).
RN   [18]
RP   VARIANT VAL-400.
RX   MEDLINE=90005402; PubMed=2792078;
RA   Tora L., Mullick A., Metzger D., Ponglikitmongkol M., Park I.,
RA   Chambon P.;
RT   "The cloned human oestrogen receptor contains a mutation which alters
RT   its hormone binding properties.";
RL   EMBO J. 8:1981-1986(1989).
RN   [19]
RP   VARIANT ALA-447.
RX   MEDLINE=92250650; PubMed=1577818.
RA   Reese J.C., Katzenellenbogen B.S.;
RT   "Characterization of a temperature sensitive mutation in the hormone
RT   binding domain of the human estrogen receptor. Studies in cell
RT   extracts and intact cells and their implications for
RT   hormone-dependent transcriptional activation.";
RL   J. Biol. Chem. 267:9868-9873(1992).
RN   [20]
RP   VARIANT GLU-364.
RX   MEDLINE=97120591; PubMed=8961262;
RA   McInerney E.M., Tsai M.-J., Katzenellenbogen B.S.,
RA   O'Malley B.W., Tsai S.Y.;
RT   "A transcriptionally active estrogen receptor mutant is a novel type
RT   of dominant negative inhibitor of estrogen action.";
RL   Mol. Endocrinol. 10:1519-1526(1996).
RN   [21]
RP   VARIANT CYS-160.
RX   MEDLINE=97338667; PubMed=9195227;
RA   Anderson T.I., Wooster R., Laake K., Collins N., Warren W., Skrede M.,
RA   Eeles R., Tveit K.M., Johnston S.R.D., Dowsett M., Olsen A.O.,
RA   Moeller P., Stratton M.R., Boerresen-Dale A.-L.;
RT   "Screening for FSP mutations in breast and ovarian cancer patients.";
RL   Hum. Mutat. 9:531-536(1997).
```

```
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC       short form; are produced by alternative splicing.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- PTM: PHOSPHORYLATED BY CYCLIN A/CDK2. PHOSPHORYLATION PROBABLY
CC       ENHANCES TRANSCRIPTIONAL ACTIVITY.
CC   -!- MISCELLANEOUS: IN THE ABSENCE OF LIGAND, STEROID HORMONE RECEPTORS
CC       ARE THOUGHT TO BE WEAKLY ASSOCIATED WITH NUCLEAR COMPONENTS;
CC       HORMONE BINDING GREATLY INCREASES RECEPTOR AFFINITY. THE
CC       HORMONE-RECEPTOR COMPLEX APPEARS TO RECOGNIZE DISCRETE DNA
CC       SEQUENCES UPSTREAM OF TRANSCRIPTIONAL START SITES.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03635; CAA27284.1; -.
DR   EMBL; M12674; AAA52390.1; -.
DR   EMBL; U47678; AAB00115.1; -.
DR   EMBL; X73067; CAA51528.1; -.
DR   EMBL; AF123500; AAD52984.1; -.
DR   EMBL; AF123494; AAD52984.1; JOINED.
```

```
  Query Match             44.0%;  Score 1233.5;  DB 1;  Length 595;
  Best Local Similarity   48.4%;  Pred. No. 1.4e-91;
  Matches  264;  Conservative   88;  Mismatches  128;  Indels   65;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSS--YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |      |  :|||   | :|: ||   | :: |  |  |  |  : |  :
Db         21 NELEPLNRP------QLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYG 74

Qy         61 NVTNLEGGPGRQ                  TTSPNVLWPTPGHLSP-LVVHRQLSHLYAE 101
                | |  ||| :                  : || :|   |  ||| |  | |    | |
Db         75 Q-TGLPYGPGSFAAAPGSNGLGGPPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLE 133

Qy        102 PQKSPWC--EAKSLELHLFVNREILKRKVSGNKCASPVIGPG-----SKRDAHFCAVCSD 154
              : | :   ||        |   :  |: |   :      |     | ::  :|||:|
Db        134 NEPSGYTVREAGPPAFYRP---NSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGGIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 269
              :| | |::| | |::: :|  |:     |: :  | |  :|
Db        251 MKGCIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMIKRSKKNS 305

Qy        259 LLDLSFPFQLVTTLLEAFPPHVLISP--PSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|  :|:|  ||:|||| :| |   |: ||:|||||  || |||:|||||:|::
Db        306 IALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRV 364

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||||:|:| ||| |||  |:|:||:||:|||::|||||:|||:|:|||::|||||::||
Db        365 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 424

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|:|:|||||:|  :::| :  :    :  :|: :||
Db        425 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 484

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
               |: ::||:|:: |||  ||| ||::|||:|| |||||||| :|||||||:|||||||
```

```
Db         485 TLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML 544

Qy         496 NAHVLRG 503
               :|| |
Db         545 DAHRLHA 552


RESULT 22
ESR1_POEGU
ID   ESR1_POEGU     STANDARD;      PRT;   587 AA.
AC   Q91250;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Poephila guttata (Zebra finch) (Taeniopygia guttata).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
OC   Estrildinae; Taeniopygia.
OX   NCBI_TaxID=59729;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hypothalamus;
RX   MEDLINE=97163508; PubMed=9010328;
RA   Jacobs E.C., Arnold A.P., Campagnoni A.T.;
RT   "Zebra finch estrogen receptor cDNA: cloning and mRNA expression.";
RL   J. Steroid Biochem. Mol. Biol. 59:135-145(1996).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- ---- ---- ---- ---- ---------- ---- -------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---- ---- ---- ---- ---------- ---- -------
DR   EMBL; L79911; AAB81108.1; -.
DR   HSSP; P03372; 1HCP.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid-binding.
FT   DOMAIN        1    176       MODULATING.
FT   DNA-BIND    177    242       NUCLEAR RECEPTOR-TYPE.
FT   ZN-FING     177    197       C4-TYPE.
FT   ZN-FING     213    237       C4-TYPE.
FT   DOMAIN      243    302       HINGE.
FT   DOMAIN      303    587       STEROID-BINDING.
SQ   SEQUENCE   587 AA;  66553 MW;  2B254168A7A9:0AB CRC64;

  Query Match          43.8%;  Score 1228.5;  DB 1;  Length 587;
  Best Local Similarity   49.3%;  Pred. No. 3.6e-91;
  Matches  261;  Conservative   86;  Mismatches  121;  Indels   61;  Gaps   14;

Qy          49 YSPAVMNYSIPS--NVTNLEGGPGRQTT-SPNVLWPTPGHLSPLVVH--RQLSHLYAE 101
               ||   ::|: |     ::| |     :   || |   |  | || : |  :|: :
Db          67 YSSTTLSYAPTSFSFGSSSLAGFHSTNSVPPSPVVFTQTAPHWSPFTHHHSQQVPYYLEN 126

Qy         102 PQKS--  ----    --PWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHF 148
                 | |             |  : |   |::   ||   |  | |   |      | ::  :
Db         127 DQGSFGMREAAPPAFYRPNSDNF--RHSI-  -RE---RMSSANEFGS--LSMESIKETRY 176

Qy         149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
               ||||:|||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db         177 CAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRK 236

Qy         209 CYEVGMVKGGSRRERGGYRLVRRQRSADFQLHCAGKAKRSGGHAPRV--RELLLD----- 261
               ||||||:| | |::| | |:::::|  :||    |:|  :   || :    |::
Db         237 CYEVGMMKGGIRKDRRGGRMVMKQKRQREEQDSFNGEASSTFLRAPTIWASPLVVKHNKKN 296

Qy         262 ----ALSPEQLVLTLLEAEPPHVLIS-RPSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
                   :|: ||:|  |||||||| |      |: || |||||  || |||:||||||:|||::
Db         297 SPALSLTAEQMVSALLEAEPPLVYSEYDPNRPFNEASMMTLLTNLADRELVHMINWAKRV 356

Qy         317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
               ||||:|:| || ||| |:|:|||:|||:|||||:|||:|:|||:|||||||||||:||
Db         357 PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 416

Qy         377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAH-LLNAVTD 435
               ||||||| :||| : || :|:|:|:|:|||||  :|  :::|  :  :  | :|: :||
Db         417 FDMLLATAARFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDYIHRVLDKITD 476

Qy         436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
               |: ::|||:| ||| ||| ||:|||:|| |||||||| ||||||||:|||||||
Db         477 TLIHLMAKSGLSLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEML 536

Qy         496 NAHVLRG- ---- -----CKSSITGSECSPAE-----DSKSKEGSQN 526
               :|| |               :| :| :  |          :|| :|  ||
Db         537 DAHRLHAPAARSAAPMEEENRSQLTTASASSHSLQSFYINSKEEENMQN 585


RESULT 23
ESR1_CHICK
ID   ESR1_CHICK     STANDARD;      PRT;   589 AA.
AC   P06212;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86247578; PubMed=3755102;
RA   Krust A., Green S., Argos P., Kumar V., Walter P., Bornert J.-M.,
RA   Chambon P.;
RT   "The chicken oestrogen receptor sequence: homology with v-erbA and
RT   the human oestrogen and glucocorticoid receptors.";
RL   EMBO J. 5:891-897(1986).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C-TERMINAL STEROID BINDING DOMAIN.
CC   -!- MISCELLANEOUS: IN THE ABSENCE OF LIGAND, STEROID HORMONE RECEPTORS
CC       ARE THOUGHT TO BE WEAKLY ASSOCIATED WITH NUCLEAR COMPONENTS;
CC       HORMONE BINDING GREATLY INCREASES RECEPTOR AFFINITY. THE
```

```
CC       HORMONE-RECEPTOR COMPLEX APPEARS TO RECOGNIZE DISCRETE DNA
CC       SEQUENCES UPSTREAM OF TRANSCRIPTIONAL START SITES
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03805; CAA27433.1; -.
DR   PIR; S07192; S07192.
DR   HSSP; P03372; 1HCP.
DR   TRANSFAC; T00264; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znt_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    178       MODULATING.
FT   DNA_BIND    179    244       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     179    199       C4-TYPE.
FT   ZN_FING     215    239       C4-TYPE.
FT   DOMAIN      245    304       HINGE.
FT   DOMAIN      305    589       STEROID-BINDING.
SQ   SEQUENCE   589 AA;  66746 MW;  1B09223307704548 CRC64;

  Query Match             43.7%;  Score 1225.5;  DB 1;  Length 589;
  Best Local Similarity   47.1%;  Pred. No. 6.3e-91;
  Matches  258;  Conservative   93;  Mismatches  127;  Indels   70;  Gaps   12;

Qy         12 SPSSYNQSQSILPLFHCSIYIPSSYVDSHHEYPAMTFYSPAVMNYS--IFSNVTNLEGGP 69
              :|:| :  :| |  |    |   :| | |     | :   || : ::|  :|  : | :|
Db         77 APTSESFGSSSLAGFHSLNNVPFSPVVPLQTAFQG--SFFIHHHSQVVFYYLENEVGSF 134

Qy         70 GRQTTSPNVIWPTPGHISPLVVHRQLSHLYAEPQKSPWCEAKSLEHTLFVNFSNHTKAEVA 129
              | :  :|      |   |   :|                                | ::: ::|
Db        134 GMREAAP------PAFYRPSSDNR----------------------------RHSIEFRMS 160

Qy        130 GNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATN 189
                  ::   :| :  :||||:|||||||||||||||||||||||||||||||||:|||||
Db        161 STNEKGSLSMESIK-ETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATN 219

Qy        190 QCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRMLKHKRQRDDGEGRNEVGSAGE 249
              |||||||||||||||||||||||:| | |::| |  :::::|  :||    |:| :
Db        220 QCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGCRMMKQKRQRDEGSRNEVGSSTE 279

Qy        250 GHAPRV--RELLLD---------ALSPEQLVLTLLEAEPPHVLIS-RPSAPFTEASMMMS 297
                :|  :    |::          :|: |||:  |||||||| |     |: || |||||
Db        280 LRAPTLWTSPLVVKHNKKNSPALSLTAEQMVSALLEAEPPIVYSEYDPNRPFNEASMMTL 339

Qy        298 LTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFA 357
              || |||:|||||:|||::||||:|:| ||| |||  |:|:||:||:|||::|||||:||
Db        340 LTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFA 399

Qy        358 PDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT 417
              |:|:|||::|||||:|||||||||| :||| : || :|:|:|:::|||||  :|  :::|
Db        400 PNLLLDRNQGKCVEGMVEIFDMLLATAARFRMMNLQGEEFVCLKSIILLNSGVYTFLSST 459

Qy        418 QDADSSRKLAH-LLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHL 476
                :  :  |   | :|: :|| |: ::|||||:| |||  ||| ||::|||:|| ||||||||
Db        460 LKSLEERDYIHRVLDKITDTLIHLMAKSGLSLQQQHRRLAQLLLILSHIRHMSNKGMEHL 519

Qy        477 LNMKCKNVVPVYDLLLEMLNAHVLR--GCKSSITGSE-------CSPAE----  ---DS 518
              |||||||||:|||||||:|| |    :|:    |          :||          :|
Db        520 YNMKCKNVVPLYDLLLEMLDAHRLHAPAARSAAPMEEENKNQLITAFASSHSLQSFYINS 579

Qy        519 KSKEGSQN 526
              | :|  ||
Db        580 KEEESMQN 587


RESULT 24
ESR1_HORSE
ID   ESR1_HORSE     STANDARD;      PRT;   594 AA.
AC   Q9TV98;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata, Craniata, Vertebrata; Euteleostomi;
OC   Mammalia, Eutheria, Perissodactyla, Equidae, Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   McDowell K.J., Adams M.H., Green M.L., Cleaver B.D., Sharp D.C.;
RL   Submitted (JAN 1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF124093; AAD17316.1; -.
DR   HSSP; P03372; 1HCP.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znt_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid binding; Phosphorylation.
FT   DOMAIN        1    184       MODULATING.
FT   DNA_BIND    185    250       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     185    205       C4-TYPE.
FT   ZN_FING     221    245       C4-TYPE.
FT   DOMAIN      251    310       HINGE.
FT   DOMAIN      311    550       STEROID-BINDING.
FT   DOMAIN       64     70       POLY-ALA.
FT   MOD_RES     104    104       PHOSPHORYLATION (BY CDK2) (BY
```

```
FT                                   SIMILARITY).
FT   MOD_RES      106    106         PHOSPHORYLATION (BY CDK2) (BY
FT                                   SIMILARITY).
FT   MOD_RES      118    118         PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES      167    167         PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   MOD_RES      536    536         PHOSPHORYLATION (BY TYR-KINASES) (BY
FT                                   SIMILARITY).
SQ   SEQUENCE   594 AA;  66103 MW;  DD36CA7C24C74B95 CRC64;

  Query Match             43.5%;  Score 1220;  DB 1;  Length 594;
  Best Local Similarity   46.0%;  Pred. No. 1.8e-90;
  Matches  267;  Conservative   99;  Mismatches  130;  Indels   84;  Gaps   19;

Qy         11 NSPSSYNCSQSTLPLEH--GSIYIPSSY--VDSHHEYPAMTFYSPAVMNYSIPSNVTNLE 66
              |   : |  |  :|||   | :|: ||   |  : |  |  | : |  : |:    : |
Db         21 NELETLNLPQFKHPSFTGFVYVFSSRPFVYTGEGAAYDFNAAAAASASVYSQ SGLA 79

Qy         67 GGPGRQTTS--PNVL--WPTPGHLSPLVVHRQLSHLYAEPQKSPW---------- 107
              ||| :  :   | |  :|    :||    || |:  || ||:
Db         80 YGPGSEAAAFGANGLGGFPPLNSVSP----SQLMLLPPPPQLSPYIHPPGQQVPYYLENE 135

Qy        108       CEA--    -KSLEHILPVNRETLKRKVSGNRCASPVTGPGSKRDAHPCAVCSD 154
                    |||        : ::    || |    ||:: :  :    | :: :||||:|
Db        136 PSGYSVCEAGPQAEYKPNADNRRQCGRERLAS--SGDKGSMAMP---SAKETPYCAVCND 190

Qy        155 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 214
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        191 YASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGM 250

Qy        215 VKGCSRERERGGYRLVRRQRSADEQLHCAGKAKRSGG---- ------------HAPRVR 256
              :| | |::| | |::: :|  |:     |: :  |                | ::
Db        251 MKGGIRKDRRGGRMLKHKRQRDD----GEGRGEVGSAGDMRAANLWPSPLIIKHSKKNS 305

Qy        257 ELLLDALSPEQLVLTLLEAEPPHVLISRPSA--PFTEASMMMSLIKLADKELVHMISWAK 314
              :|  :|: |:::  ||:|||| || |   |  || |||||  || |||:|||||:|||
Db        306 PVL--SLTAEQMISALLDAEPP-VLYSEYDATRPFNEASMMGLLTNLADRELVHMINWAK 362

Qy        315 KIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGIL 374
              ::|||:||| ||| ||| |:|:||:|||:|::|||||:|||:|:|||::|||||||::
Db        363 RVPGFVDLSLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMV 422

Qy        375 EIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAV 433
              |||||||||:|| | || :|:::|:|::|||||  :|  :::| :  :    :  :: :
Db        423 EIFDMLLATSSRLRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKM 482

Qy        434 TDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLE 493
              || |: ::||:|:: ||   ||| ||::||:||:|| |||||||| :|||||||:||||||
Db        483 TDTLIHLMAKAGLTLQQH-RRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLE 541

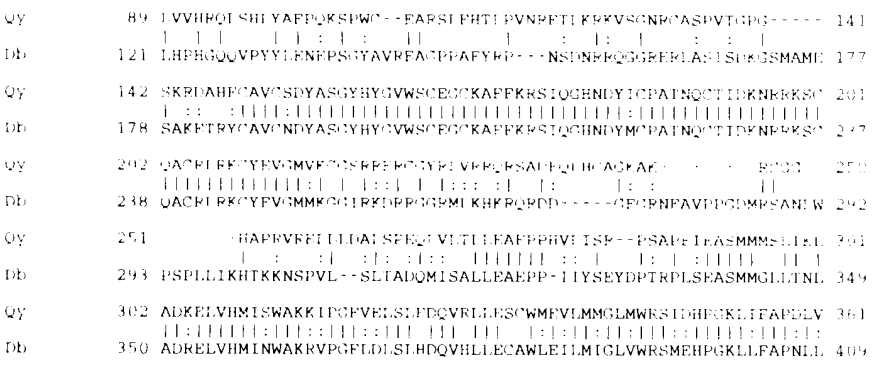
Qy        494 MLNAHVLRGCKSSITGSECSPAEDSKSKE-   CSQNPQS 529
              ||:|| |     :      :| |::   :    || :| |
Db        542 MLDAHRLH--  APANHGGAPMEETNQSQLATTGSTSPHS 577


RESULT 25
ESR1_PIG
ID   ESR1_PIG        STANDARD;       PRT;   595 AA.
AC   Q29040;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Uterus;
RX   MEDLINE=95080454; PubMed=7988744;
RA   Boockenkamp D., Jungblut P.W., Thole H.H.;
RT   "The C-terminal half of the porcine estradiol receptor contains no
RT   post-translational modification: determination of the primary
RT   structure.";
RL   Mol. Cell. Endocrinol. 104:163-172(1994).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- PTM: PHOSPHORYLATED BY CYCLIN A/CDK2 (BY SIMILARITY).
CC   -!- MISCELLANEOUS: IN THE ABSENCE OF LIGAND, STEROID HORMONE RECEPTORS
CC       ARE THOUGHT TO BE WEAKLY ASSOCIATED WITH NUCLEAR COMPONENTS;
CC       HORMONE BINDING GREATLY INCREASES RECEPTOR AFFINITY. THE
CC       HORMONE-RECEPTOR COMPLEX APPEARS TO RECOGNIZE DISCRETE DNA
CC       SEQUENCES UPSTREAM OF TRANSCRIPTIONAL START SITES.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; Z37167; CAA85524.1; -.
DR   HSSP; P03372; 1HCP.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znt_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding; Phosphorylation.
FT   DOMAIN        1    184         MODULATING.
FT   DNA_BIND    185    250         NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     185    205         C4-TYPE.
FT   ZN_FING     221    245         C4 TYPE.
FT   DOMAIN      251    310         HINGE.
FT   DOMAIN      311    551         STEROID-BINDING.
FT   DOMAIN       64     70         POLY-ALA.
FT   MOD_RES     104    104         PHOSPHORYLATION (BY CDK2) (BY
FT                                  SIMILARITY).
FT   MOD_RES     106    106         PHOSPHORYLATION (BY CDK2) (BY
FT                                  SIMILARITY).
FT   MOD_RES     118    118         PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     167    167         PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   MOD_RES     537    537         PHOSPHORYLATION (BY TYR-KINASES) (BY
FT                                  SIMILARITY).
SQ   SEQUENCE   595 AA;  66461 MW;  0EC24DC3085BF001 CRC64;

  Query Match             44.5%;  Score 1219.5;  DB 1;  Length 595;
  Best Local Similarity   46.0%;  Pred. No. 1.9e-90;
  Matches  268;  Conservative   89;  Mismatches  126;  Indels   99;  Gaps   19;

Qy          5 NSPSSLNSPSSYNCSQSTLPLEH-GSIYIPSSYVDSHHEYPAMTFYSPAVMNY     -- 56
              |    || |       |  :|||   | :|: ||                 ||| ||
Db         21 NELEPLNRP- - --QLKIPLERPLGEVYVDSS---  --- -  - KPAVYNYPEGAAY 60

Qy         57 --------SIP-SNVTNLEGGPGRQ--------  -----   --TISPNVLWPIPGHLSP- 88
                      || |   : |  ||| :                    : || ||   |  |||
Db         61 DFNAAAAASAPVYGQSGLAYGPGSEAAAFGANGLGGFQPLNSVSPSPLVLLHPPPQLSPF 120
```



```
RESULT 26
ESR1_MESAU
ID   ESR1_MESAU     STANDARD;      PRT;   595 AA.
AC   Q9QZJ5; Q9QZG6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha)
GN   ESR1 OR NR3A1 OR ESR
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=20197937; PubMed=10731637;
RA   Bhat H.K., Vadgama J.V.;
RT   "Hamster estrogen receptor cDNA, cloning and mRNA expression.";
RL   J. Steroid Biochem. Mol. Biol. 72:47-53(2000).
RN   [2]
RP   SEQUENCE OF 98-291 FROM N.A.
RA   Jones J.E., Carpenter C.D., Lubbers L.S., Petersen S.L., Wade G.N.;
RT   "Return of lordosis after food deprivation and refeeding in Syrian
RT   hamsters.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF181277; AAD53956.1; -.
DR   EMBL; AF184276; AAF02721.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid-binding; Phosphorylation.
FT   DOMAIN        1    184       MODULATING.
FT   DNA_BIND    185    250       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     185    205       C4-TYPE.
FT   ZN_FING     221    245       C4-TYPE.
FT   DOMAIN      251    310       HINGE.
FT   DOMAIN      311    551       STEROID-BINDING.
FT   DOMAIN      112    115       POLY-PRO.
FT   DOMAIN      166    169       POLY-SER.
FT   MOD_RES     103    103       PHOSPHORYLATION (BY CDK2) (BY
FT                                SIMILARITY).
FT   MOD_RES     105    105       PHOSPHORYLATION (BY CDK2) (BY
FT                                SIMILARITY).
FT   MOD_RES     118    118       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     167    167       PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   MOD_RES     537    537       PHOSPHORYLATION (BY TYR KINASES) (BY
FT                                SIMILARITY).
FT   CONFLICT     98     99       QL -> PF (IN REF. 2).
FT   CONFLICT    144    144       S -> T (IN REF. 2).
FT   CONFLICT    198    199       GV -> EVEQI (IN REF. 2).
FT   CONFLICT    290    291       NL -> KG (IN REF. 2).
SQ   SEQUENCE   595 AA;  66833 MW;  E4D803B4FFDB257E CRC64;

  Query Match          43.5%;  Score 1219;  DB 1;  Length 595;
  Best Local Similarity   46.7%;  Pred. No. 2.1e-90;
  Matches  262;  Conservative   93;  Mismatches  134;  Indels   72;  Gaps   15;

Qy         23 LPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNY----------SIPSNV---TNLEGG 68
              ||||     :   ||||           ||: ||          : |: |   | :  |
Db         33 MPLERA---LSEVYVDSS---------KPAMFNYPEGAAYEFNAATAPAPVYGQTGIAYG 80

Qy         69 PGRQTT-----------------SPN---VLWPTPGHLSP-LVVHRQLSHLYAEPQKSPWC 108
              | : |                 ||:   :| | |  ||| |  | |    | | : | :
Db         81 SGSEATAPGSNSLGLFPQLNSVSPSPLMLLHPPPPQLSPFLHPHGQQVPYYLENEPSAYA 140

Qy        109 FARSLEHTIPVNRFTLKRKVSGNRCASPVTGPGSK-----RDAHFCAVCSDYASGYHYGV 163
                |           :  |: ||    |  :  ||       ::  :||||:||||||||
Db        141  VRVSGPLAFYEPNSENERQSGREELSSSSEEIS MAMELVFETFYGAVCNDYASGYHYGV 199

Qy        164 WSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDR 223
              ||||||||||||||||||||:|||||||||||||||||||||||||||||:| | |:||
Db        200 WSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDR 259

Qy        224 CGYPLVPQPSADP-----QLHCAGKAKRSG--------GHAPFVRELLLDALSPEQLVL 270
              | |::: :|  |:     :  :|  : :              | :    |  :|: :|:|
Db        260 RGGRMLKHKRQRDDLEGRNDMGPSGDMRATNIWPSPLVIKHTKKNSPAL--SLTADQMVS 317

Qy        271 TLLEAEPPHVLIS-RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVR 329
              ||:|||| :     || ||:|||||  || |||:|||||:|||::||| :|:| |||
```

```
Db        318 ALLDAEPPLIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRVPGFGDLNLHDQVH 377

Qy        330 LLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRE 389
              ||| |:|:||:|:|||::|||||:|||:|:|||::|||||||::||||||||||:|||
Db        378 LLECAWLEILMIGLIWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSARFRM 437

Qy        390 LKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGLSS 449
              : || :|:::|:::||||| :|  :::| :  :    :  :|: :|| |: ::||:|::
Db        438 MDLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTL 497

Qy        450 QQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT 509
              ||| |||:|||:|| |||||||| |||||||| ||||||||:|| |    : ::
Db        498 QQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPFYDLLLEMLDAHRLH -TPVS 554

Qy        510 GSECSPAEDSKSKEGSQNPQS 529
                 || | |:|:  : | |
Db        555 RMGVSPEEPSQSQLETTNSTS 575


RESULT 27
ESR1_MOUSE
ID   ESR1_MOUSE     STANDARD;      PRT;   599 AA.
AC   P19785; Q9QY51; Q9QY52;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR OR ESTR OR ESTRA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=91042558; PubMed=2484714;
RA   White R., Lees J.A., Needham M., Ham J., Parker M.;
RT   "Structural organization and expression of the mouse estrogen
RT   receptor.";
RL   Mol. Endocrinol. 1:735-744(1987).
RN   [2]
RP   SEQUENCE OF 269-599 FROM N.A.
RC   STRAIN=SJL/J and B10.S/J; TISSUE=Spleen;
RA   Ma R.Z., Teuscher C.;
RT   "Screening for candidate genes of mouse autoimmune diseases.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CARBOHYDRATE-LINKAGE SITE=SER-575.
RX   MEDLINE=97158020; PubMed=8999954;
RA   Jiang M.S., Hart G.W.;
RT   "A subpopulation of estrogen receptors are modified by O-linked
RT   N-acetylglucosamine.";
RL   J. Biol. Chem. 272:2421-2428(1997).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES=SER-10; THR-50 AND SER-575.
RX   MEDLINE=21124487; PubMed=11226831;
RA   Cheng X., Hart G.W.;
RT   "Glycosylation of the murine estrogen receptor-alpha.";
RL   J. Steroid Biochem. Mol. Biol. 75:147-158(2000).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- PTM: PHOSPHORYLATED BY CYCLIN A/CDK2 (BY SIMILARITY).
CC   -!- MISCELLANEOUS: IN THE ABSENCE OF LIGAND, STEROID HORMONE RECEPTORS
CC       ARE THOUGHT TO BE WEAKLY ASSOCIATED WITH NUCLEAR COMPONENTS;
CC       HORMONE BINDING GREATLY INCREASES RECEPTOR AFFINITY. THE
CC       HORMONE-RECEPTOR COMPLEX APPEARS TO RECOGNIZE DISCRETE DNA
CC       SEQUENCES UPSTREAM OF TRANSCRIPTIONAL START SITES.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- -------- ---------- - -- . - - -- ----- - -
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -- ------------------------------------ - -- - ---------- -- ----
DR   EMBL; M38651; AAA37580.1; -.
DR   EMBL; AF128221; AAF22562.1; -.
DR   EMBL; AF128220; AAF22561.1; -.
DR   PIR; A40061; A40061.
DR   HSSP; P03372; 1HCP.
DR   TRANSFAC; T00259; -.
DR   GlycoSuiteDB; P19785; -.
DR   MGD; MGI:1352467; Esr1.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   Prodom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid binding; Phosphorylation; Glycoprotein.
FT   DOMAIN        1    188       MODULATING.
FT   DNA_BIND    189    254       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING     189    209       C4-TYPE.
FT   ZN_FING     225    249       C4-TYPE.
FT   DOMAIN      255    314       HINGE.
FT   DOMAIN      315    599       STEROID-BINDING.
FT   DOMAIN       64     72       POLY-ALA.
FT   MOD_RES     108    108       PHOSPHORYLATION (BY CDK2) (BY
FT                                SIMILARITY).
FT   MOD_RES     110    110       PHOSPHORYLATION (BY CDK2) (BY
FT                                SIMILARITY).
FT   MOD_RES     122    122       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     171    171       PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   MOD_RES     541    541       PHOSPHORYLATION (BY TYR KINASES) (BY
FT                                SIMILARITY).
FT   CARBOHYD     10     10       O-LINKED (GLCNAC).
FT   CARBOHYD     50     50       O-LINKED (GLCNAC).
FT   CARBOHYD    575    575       O-LINKED (GLCNAC).
FT                                /FTId=CAR_000137.
FT   VARIANT     591    591       E -> Q (IN STRAIN SJL/J).
FT   CONFLICT    269    269       L -> M (IN REF. 2).
SQ   SEQUENCE   599 AA;  66955 MW;  05F5E2FC21CC0A8B CRC64;

  Query Match             43.4%;  Score 1218.5;  DB 1;  Length 599;
  Best Local Similarity   45.8%;  Pred. No. 2.3e-90;
  Matches  263;  Conservative   98;  Mismatches  152;  Indels   61;  Gaps   14;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH-GSIYIPSS-YVDSHHEYPAMTFYSPAVMNYSIPS 60
              |    || |       | :|:|   | :|: :|   | :: | | | | : |    : :
Db         21 NELEPLNRP-----QLKMPMERALGEVYVDNSKPTVFNYPEGAAYEFNAAAAAAAASA 74

Qy         61 NVINLEG---GPGRQ---------- ------TTSPNVIWPTPGHLSP-LVVHRQLSHL 98
              |    |   ||| :                : || :|   |  ||| | | |
Db         75 PVYGQSGIAYGPGSEAAAFSANSLGAFPQLNSVSPSPLMLLHPPPQLSPFLHPHGQQVPY 134

Qy         99 YAEPQKSPWCEARSLEHTLP---VNRETLKRKVSGNRCASPVTGPG - -SKRDAHFCA 150
              | | : |   | ::  | |       :  |: :|   |      |   | ::  :||
Db        135 YLENEPS----AYAVRDTGPPAFYRSNSDNRRQNGRERLSSSNEKGNMIMESAKETRYCA 190
```

```
Qy        151 VCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCY 210
              ||:|||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        191 VCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCY 250

Qy        211 EVGMVKGGSRRERGGYRLVRRQRSADE-----QLHCAGKAKRSG--------GHAPRVRE 257
              ||||:| | 1::| | |:::.:|  |:     ::  :|  : :          | :
Db        251 EVGMMKGGIRKDRRGGRMLKHKRQRDDLEGRNEMGASGDMRAANLWPSPLVIKHTKKNSP 310

Qy        258 LITDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKI 316
              |  :|: :|:|  ||:|||| :     || ||:|||||  || |||:||||||:|||::
Db        311 AL--SLTADQMVSALLDAEPPMIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRV 368

Qy        317 PGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEI 376
              ||| :|:| ||| |||  |:|:||:||:|||::|||||:|||:|:|||::|||||::||
Db        369 PGFGDLNLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEI 428

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTD 435
              |||||||:|||| : || :|:::|:|::|||| :|  :::| :  .     :  :|: :||
Db        429 FDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITD 488

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
              |: ::||:|:: ||| |||| ||::|||:|| |||||||| ||||||||:|||||||||
Db        489 TLIHLMAKAGLTLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEML 548

Qy        496 NAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQS 529
              :|| |    | :      |::   :   | : |
Db        549 DAHRLHAPASRMGVPPEEPSQTQLATTSSTSAHS 582


RESULT 28
ESR1_RAT
ID   ESR1_RAT        STANDARD;      PRT:   600 AA.
AC   P06211;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR OR ESTR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RA   Muramatsu M.;
RL   Submitted (MAR-1987) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87174780; PubMed=3031601;
RA   Koike S., Sakai M.;
RT   "Molecular cloning and characterization of rat estrogen receptor
RT   cDNA.";
RL   Nucleic Acids Res. 15:2499-2513(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Uterus;
RA   Maqqi A.M.A.;
RL   Submitted (JUN-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- PTM: PHOSPHORYLATED BY CYCLIN A/CDK2 (BY SIMILARITY).
CC   -!- MISCELLANEOUS: IN THE ABSENCE OF LIGAND, STEROID HORMONE RECEPTORS
CC       ARE THOUGHT TO BE WEAKLY ASSOCIATED WITH NUCLEAR COMPONENTS;
CC       HORMONE BINDING GREATLY INCREASES RECEPTOR AFFINITY. THE
CC       HORMONE-RECEPTOR COMPLEX APPEARS TO RECOGNIZE DISCRETE DNA
CC       SEQUENCES UPSTREAM OF TRANSCRIPTIONAL START SITES.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------- - -
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y00102; CAA68287.1; -.
DR   EMBL; X61098; CAA43411.1; -.
DR   PIR; S07379; S07379.
DR   HSSP; P03372; 1HCP.
DR   TRANSFAC; T00258; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding; Phosphorylation.
FT   DOMAIN       1    189       MODULATING.
FT   DNA_BIND     190  255       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      190  210       C4 TYPE.
FT   ZN_FING      226  250       C4 TYPE.
FT   DOMAIN       256  315       HINGE.
FT   DOMAIN       316  600       STEROID-BINDING.
FT   DOMAIN        64   71       POLY-ALA.
FT   DOMAIN       171  174       POLY-SER.
FT   MOD_RES      109  109       PHOSPHORYLATION (BY CDK2) (BY
FT                               SIMILARITY).
FT   MOD_RES      111  111       PHOSPHORYLATION (BY CDK2) (BY
FT                               SIMILARITY).
FT   MOD_RES      123  123       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES      172  172       PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   MOD_RES      542  542       PHOSPHORYLATION (BY TYR-KINASES) (BY
FT                               SIMILARITY).
FT   CONFLICT     488  488       N -> T (IN REF. 3).
SQ   SEQUENCE   600 AA;  67030 MW;  C9C7D8CACE0E57D8 CRC64;

  Query Match             43.4%;  Score 1216.5;  DB 1;  Length 600;
  Best Local Similarity   45.7%;  Pred. No. 3.4e-90;
  Matches  262;  Conservative   99;  Mismatches  135;  Indels   77;  Gaps   17;

Qy          5 NSPSSLNSPSSYNCSQSILPLEH--GSIYIPSSYVDSHHEYPAMTFY------SPAVMNY 56
              |    || |      | :|:|   | :|: :|   :   ||   |       : |
Db         21 NELEPLNRP------QLKMPMERALGEVYVDNS-KPAVFNYPEGAAYEFNAAAAAAAAGA 73

Qy         57 SIP----SNVTNLEGGPGRQ-----------------TTSPNVLWPTPGHLSPLV   V 91
              | |    |::|    |||  :                  : || :|   | |:|| :
Db         74 SAPVYGQSSITY --GPGSEAAAFGANSLGAFPQLNSVSPSPLMLLHPPPHVSPFLHPHG 130

Qy         92 HPQISHLYAFPQKSPWCPAPSLEHILP---VNPELLGRRKVSGNRCASPVTGPG-----SK 143
              |:   :|  ||          | ::  | |        :  |: :|    |  :  |      |
Db        131 HQVPYYLENEP------SAYAVRDTGPPAFYRSNSDNRRQNGKERLSSSSEKGNMIMESA 184

Qy        144 RDAHFGAVGSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQA 203
              ::  :|||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db        185 KETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQA 244

Qy        204 CRLRKCYEVGMVKGGSRRERGGYRLVRRQRSADE-----QLHCAGKAKRSG--------G 250
              |||||||||||:| | |::| | |::: :|  |:     ::  :|  : :
Db        245 CRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDLEGRNEMGASGDMRAANLWPSPLVIK 304
```

```
Qy          251 HAPRVRELLLLDALSPEQLVLTLLEAEPPHVLIS-RPSAPFTEASMMMSLTKLADKELVHM 309
                |  :    |  :|: :|:|  ||:|||| :     || ||:|||||  || |||:|||||
Db          305 HTKKNSPAL  SLTADQMVSALLDAEPPLIYSEYDPSRPFSEASMMGLLTNLADRELVHM 362

Qy          310 ISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKC 369
                |:|||::||| :|:|  ||| ||||  |:|:||:||:|||::|||||:|||:|:|||::|||
Db          363 INWAKRVPGFGDLNLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKC 422

Qy          370 VEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAH 428
                |||::|||||||||-|||| : || :|:|:|:|:|||| :|  :::| :  :    :
Db          423 VEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHR 482

Qy          429 LLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVY 488
                :|: : | |: ::||:|:: ||| ||| ||::|||:|| |||||||| ||||||||:|
Db          483 VLDKINDTLIHLMAKAGLTLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLY 542

Qy          489 DLLLEMLNAHVLRGCKSSITGSECSPAEDSKSK 521
                |||||||:|| |    | :      | | |:|:
Db          543 DLLLEMLDAHRLHAPASRM-  GVPPEEPSQSQ 572


RESULT 29
ESR1_ORENI
ID   ESR1_ORENI     STANDARD;      PRT;   585 AA.
AC   Q9YH33;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidei;
OC   Cichlidae; Oreochromis.
OX   NCBI_TaxID=8128;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chang X.T., Kobayashi T., Todo T., Yoshiura Y., Ikeuchi T.,
RA   Kajiura H., Nakamura M., Nagahama Y.;
RT   "cDNA sequence of Tilapia type alpha estrogen receptor.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC    !  FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC    !  SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER
CC       BETA (BY SIMILARITY).
CC    !  SUBCELLULAR LOCATION: Nuclear.
CC    !  DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C-TERMINAL STEROID BINDING DOMAIN.
CC    !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; U75604; AAD00245.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; transcription regulation; DNA binding; Nuclear protein;
KW   zinc-finger; steroid-binding.
FT   DOMAIN          1     138       MODULATING (BY SIMILARITY).
FT   DNA_BIND      139     204       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING       139     159       C4-TYPE.
FT   ZN_FING       175     199       C4-TYPE.
FT   DOMAIN        205     266       HINGE.
FT   DOMAIN        267     585       STEROID-BINDING (BY SIMILARITY).
FT   DOMAIN        250     255       POLY-SER.
FT   DOMAIN        463     467       POLY-LEU.
SQ   SEQUENCE   585 AA;  64165 MW;  36A96BCB668707B0 CRC64;

  Query Match             44.3%;  Score 1214;  DB 1;  Length 585;
  Best Local Similarity   49.4%;  Pred. No. 5.2e-90;
  Matches  272;  Conservative   68;  Mismatches  147;  Indels   74;  Gaps   16;

Qy           36 YVDSHHEYPAMT- ---FYSPAVMN-YSIP  --   -  SNVTNLEGGPGRQTISPNVLW 80
                ::::  ::| | |      || :    || |          :: :|  ||    ||| |
Db           17 FLEGTYDYAAPTPAPTPLYSHSTTGCYSAPLDAHGPLSDGSLQSLGSGP----TSPLVFV 72

Qy           81 PTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHILPVNRE----TLKRKVSGNRCASP 136
                |:   ||| :     || | |   :|  :    |  || ||   |     |     :
Db           73 PSSPRLSPFM--HPPSHHYLETTSTPVYRS---SHQ-PVPREDQCGTRDEAYSVGELGAG 126

Qy          137 VTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKN 196
                |    :|  ||||||||||||||||||||||||||||||||||||||:||||||||||
Db          127 AGGFEMTKDTRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKN 186

Qy          197 RRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVKRQRSA----------------DEQ 238
                ||||||||||||||||||:| | |::| |  | | :| |                  | :
Db          187 RRKSCQACRLRKCYEVGMMKGGMRKDR-GRVLRREKRRACDRDKPAKDLPHTRASPQDGR 245

Qy          239 LHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP--SAPFTEASMMM 296
                      :  |||          |: : |:|::| |  |||| :| ||   | |:|| ::|
Db          246 KRAMSSSSTSGGGG----PSSLNNMPPDQVTLLLQCAEPP-ILSSPQKMSRPYTFVTIMT 300

Qy          297 SLIKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIF 356
                 || :|||||||||:||||:|||:::||| ||| |||| |:||||:||:||||  ||||||
Db          301 LITSMADKFLVHMITWAKKIPGFIQLSLHDQVILLESSWLEVLMIGLIWRSIQCPGKLIF 360

Qy          357 APDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTA 416
                | ||:|||:|| |||:  |||||||||| |||| ||| :|:|:|:||:|||||   :  |
Db          361 AQDLILDRNEGTCVEGMAEIFDMLLATASRFRVLKLKPEEFVCLKAIILLNSGAFSFCTG 420

Qy          417 TQD-ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEH 475
                | :    |  : |:|: :|||||:: |:  | |:|||| | | ||:|||||:|| ||||||
Db          421 IMEPLHDSAAVQRMLDTITDALIFRISHLGCSAQQQSRRQAQLLLLLSHIRHMSNKGMEH 480

Qy          476 LLNMKCKNVVPVYDLLLEMLNAHVL - --       RGCKSSITGSECSPA-      -ED 517
                | :|||| ||:|||||||:|| :            :| : | | |  | :        |
Db          481 LYSMKCKNKVPLYDLLLEMLDAHRIHAPVKPPQSWSQGDRDSPTASSTSSSGGGGGDDEG 540

Qy          518 SKSKEGSQNPQ 528
                : |   |  ||
Db          541 ASSAGSSSGPQ 551


RESULT 30
ESR1_XENLA
ID   ESR1_XENLA     STANDARD;      PRT;   586 AA.
AC   P81559;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
```

```
OC   Xenopodinae; Xenopus
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90331927; PubMed=2274894;
RA   Weiler I.J., Lew D., Shapiro D.J.;
RT   "The Xenopus laevis estrogen receptor: sequence homology with human
RT   and avian receptors and identification of multiple estrogen receptor
RT   messenger ribonucleic acids.";
RL   Mol. Endocrinol. 1:355-362(1987).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED
CC       IN THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT
CC       CELLULAR PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch, not work,
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------- --
DR   EMBL; L20735; -; NOT_ANNOTATED_CDS.
DR   HSSP; P03372; 1HCQ.
DR   TRANSFAC; T00263; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    179       MODULATING.
FT   DNA_BIND    180    245       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     180    200       C4 TYPE.
FT   ZN_FING     216    240       C4-TYPE.
FT   DOMAIN      246    302       HINGE.
FT   DOMAIN      303    586       STEROID-BINDING
SQ   SEQUENCE   586 AA;  66080 MW;  0EDC77EB0D6F08BF CRC64;

  Query Match             43.3%;  Score 1214;  DB 1;  Length 586;
  Best Local Similarity   48.6%;  Pred. No. 5.2e-90;
  Matches 255;  Conservative   82;  Mismatches  139;  Indels   5;  Gaps   2;

Qy         36 YVDSHH----EYP----------AMTFYSPAVMNYSIPSNVINLEGGPGRQT---SISP 76
              ||:::      ||          |   || | ::|:  |         |  |      ||
Db         43 YVENNRTGIFNYPEGTTYDFAAAAAPVYSSASLSYAASSETPGSSSLTGLHTLNNVPPSP 102

Qy         77 NVIWPTPCHLSPIVVHP-QISHLYAPPLFSFWCEAPSLEHTLFVNEELLPEPVSDHSFAS 135
              |       ||| : |  |    | | ::  :    :   |       :  |:||   |
Db        103 VVFIAFLPQISPETHHHCQQVPYYLESPCCIFAVPEAAPPIF-YRSSSDNRPQSGPEPMS 161

Qy        136 PVTGPG-----SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
                  |     | ::  :|||||||||||||||||||||||||||||||||||||||||||||||
Db        162 SANDKGPPSMESTKFTRYCAVCSDYASGYHYGVWSCPGCKAFFKRSIQGHNDYMCPATNQ 221

Qy        191 CTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQKEEQ-----EQKNDVD 245
              ||||||||||||||||||||||:| | |::| | |:: :|  :||     : |
Db        222 CTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQKEEQ-----EQKNDVD 276

Qy        246 HA----------PRVRELLLD---ALSPEQLVLTLLEAEPPHVLISRPSA-PFTEASMMM 296
              :           | |: : |    :|: |||:  |:|||| | |      |  | :|||||
Db        277 PSEIRTASIWVNPSVKSMKLSPVLSLTAEQLISALMEAEAPIVYSEHDSIKPLSEASMMI 336

Qy        297 SLIFLADFELVHMISWAKEIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIF 356
              || |||:||||||:|||:|||||:| ||| |||  |:|:|||:||:|||:|||||| |
Db        337 LLINLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMVGLIWRSVEHPGKLSF 396

Qy        357 APDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTA 416
              |||:|||:::|||||::||||||: | :||| :::|: :|:::|:|:::|||| :|  :::
Db        397 APNLLLDRNQGRCVEGLVEIFDMLVHTAIRFRMMNLQGEEFVCLKSIILLNSGVYTFLSS 456

Qy        417 TQDADSSRKLAH-LLNAVIDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEH 475
              | ::     | | :|: : | ||  :||||:| |||  ||| ||::|||:|| |||||||
Db        457 TLPSLEDIDLIHFILDKIIDTLVHFMAKSGLSLQQQQRRLAQLLLILSHIRHMSNKGMEH 516

Qy        476 LLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKS 520
              | :|||||||:|||||||:|| :   |  | |    |||:|
Db        517 LYSMKCKNVVPLYDLLLEMLDAHRLHIDRDKTITQPT--  -EDSRS 557


RESULT 31
ESR1_BRARE
ID   ESR1_BRARE     STANDARD;      PRT:   569 AA.
AC   P57717;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Morimoto J., Miyasho T., Iwano H., Teraoka H., Hiraga T., Yokota H.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------- --
DR   EMBL; AB037185; BAB16893.1; -.
DR   HSSP; P03372; 1HCP.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
```

```
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid-binding.
FT   DOMAIN          1     151       MODULATING (BY SIMILARITY).
FT   DNA_BIND      152     217       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING       152     172       C4-TYPE.
FT   ZN_FING       188     212       C4-TYPE.
FT   DOMAIN        218     278       HINGE.
FT   DOMAIN        279     569       STEROID BINDING (BY SIMILARITY).
FT   DOMAIN        284     288       POLY-LEU.
FT   DOMAIN        475     479       POLY-LEU.
SQ   SEQUENCE   569 AA;  62826 MW;  4D2B57599P6D9132 CRC64;

  Query Match           42.9%;  Score 1204;  DB 1;  Length 569;
  Best Local Similarity 48.7%;  Pred. No. 3.2e-89;
  Matches  275;  Conservative   71;  Mismatches  141;  Indels   88;  Gaps   18;

Qy         26 EHGSIYIPSS--YVDSHHEYP----AMTFYSPA ---VMNYSIP---- SNVTNLEGGPG 70
              ||  :  | ||  |:|  :|||     |    |||     |  |  |         ::  | ||
Db          6 EHSAGGISSSVNYLDGAYEYPNPTQAFGTSSPAEPASVGYYPAPPDPHEEHLQTLGGG- 63

Qy         71 RQTTSPNVLWPTPGHLSPLVVH--------RQLSHLYAEPQKSP------- ---- - W 107
               ::|| :  |:   ||| : |           |:|: | :   |
Db         64 --SSSPLMFAPSSPQLSPYLSHHGGHHTTPHQVSY-YLDSSSSTVYRSSVVSSQQAAVGL 120

Qy        108 CEARSLEHTLPVNRETLKRKVSGNRCASPVTGPQSKRDAHFCAVCSDYASGYHYGVWSCE 167
              ||    |     :|: |     :|:| |    |  | :: ||||||||||||||||||
Db        121 CE----ELCSATDRQELK---YTGSRAAG---GFDSGKETRFCAVCSDYASGYHYGVWSCE 170

Qy        168 GCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGGYR 227
              ||||||||||||||||:||||||||||:|||||||||||||||||||:| | |::| | |
Db        171 GCKAFFKRSIQGHNDYVCPATNQCTIDRNKRKSCQACRLRKCYEVGMMKGGIRKDR-GGR 229

Qy        228 LVRRQR --   ---  -SADEQLHCAG --------KAKRSGGHAPRVRELLLDALSPEQL 268
              |||:|          |: :|   ||         : ||||     :         : |:|:
Db        230 SVRRFRRRSSNFTPDKSSSDQCSRACVPTTGPQDKPKKRSCCVVSTL- ----CMPPDQV 284

Qy        269 VLTLLEAEPPHVLI SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQ 327
              :| || |||| |    : | :|:|| :||  || :|||||||||:||||:||| :||| ||
Db        285 LLLLIGAEPFAVCSRQKHSRPYTFTIMMSLTTNMADKELVHMIAWAKKVPGFLDLTLHDQ 344

Qy        328 VRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRF 387
              |:|||| |:||||:||:|||  |||||||  ||:||| ||:|||: |||||||||  :||
Db        344 VQLLESSWLEVLMIGLIWRSIDHPGKLIFAQDLILDRSEGECVEGMAEIFDMLLATVARF 403

Qy        388 RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAH-LLNAVTDALVWVIAKSGI 446
              | |||: :|:::|:||:||  :   ::  :     :  :|: :|||:: |:|||
Db        404 RSLKLKLEEFVCLKAIILLNSGAFSFCSSPVEPLMDNFMVQCMLDNITDALIYCISKSGA 463

Qy        447 SSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH--    --V 499
              | | || | | ||:|||:|| |||||||  ||||| ||:|||||||:|     |
Db        464 SLQLQSRRQAQLLLLLSHIRHMSNKGMEHLYRMKCKNIVPLYDLLLEMLDAQRFQSSGKV 523

Qy        500 LRGCKSSITGSECSPAEDSKSKEGS 524
              |   |     :|   | |   |
Db        524 QRVWSQSEKNPPSTPTTSSSSSNNS 548


RESULT 32
ESR1_SPAAU
ID   ESR1_SPAAU     STANDARD;      PRT;   579 AA.
AC   Q9PVZ9; Q9PWP6; O42174;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Sparus aurata (Gilthead sea bream).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
OC   Sparidae; Sparus.
OX   NCBI_TaxID=8175;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=99304495; PubMed=10376207;
RA   Munoz-Cueto J.A., Burzawa-Gerard E., Kah O., Valotaire Y., Pakdel F.;
RT   "Cloning and sequencing of the gilthead sea bream estrogen receptor
RT   cDNA.";
RL   DNA Seq. 10:75-84(1999).
RN   [2]
RP   SEQUENCE OF 52-579 FROM N.A.
RC   TISSUE=Liver;
RX   PubMed=10927619;
RA   Socorro S., Power D.M., Olsson P.-E., Canario A.V.M.;
RT   "Two estrogen receptors expressed in the teleost fish, Sparus aurata:
RT   cDNA cloning, characterization and tissue distribution.";
RL   J. Endocrinol. 166:293-306(2000).
RN   [3]
RP   SEQUENCE OF 208-484 FROM N.A.
RC   TISSUE=Liver;
RA   Socorro S., Power D.M., Canario A.V.M.;
RT   "cDNA cloning and tissue expression of the gilthead sea-bream, Sparus
RT   aurata, estrogen receptor.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   - ---- -- ------- ------------------------ ---------- ----- ----- -- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -- ------------------------ - ---- -- . . - --- -- ------
DR   EMBL; AJ006039; CAB51479.1; -.
DR   EMBL; AF136979; AAD31042.1; -.
DR   EMBL; AF013104; AAB82428.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000546; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   Prodom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN          1     142       MODULATING (BY SIMILARITY).
FT   DNA_BIND      143     208       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING       143     163       C4-TYPE.
FT   ZN_FING       179     203       C4 TYPE.
FT   DOMAIN        209     270       HINGE.
FT   DOMAIN        271     579       STEROID-BINDING (BY SIMILARITY).
FT   DOMAIN        259     263       POLY-GLY.
FT   DOMAIN        533     537       POLY-SER.
FT   DOMAIN        538     543       POLY-GLY.
FT   CONFLICT       83      86       QPAN -> HPPT (IN REF. 2).
FT   CONFLICT       98      98       Y -> YRS (IN REF. 2).
FT   CONFLICT      480     480       K -> N (IN REF. 3).
SQ   SEQUENCE   579 AA;  63361 MW;  7FB13C7AEC56371F CRC64;
```

```
  Query Match             42.9%;  Score 1204;  DB 1;  Length 579;
  Best Local Similarity   48.4%;  Pred. No. 3.3e-89;
  Matches  261;  Conservative   82;  Mismatches  128;  Indels   68;  Gaps   14;

Qy          36 YVDSHHEYPAMT-------------FYSPAVMNYSIPS--NVTNIFGGPGPQTISPNVLW 80
               :::  ::| | |              :|| :  :  || :: :|  ||     || |
Db          17 FLEGTYDYAAPTPAPTPLYSHSTPGYYSAPLDAHGPPSDGSLQSLGSGP----NSPLVFV 72

Qy          81 PTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETL------KRKVSGNRCA 134
               |:  |||| :  :| | |  :|     | :|:  |:||              | :
Db          73 PSSPHLSPFM--QPANHHYLETTSIPIYSVPSSQHS--VSREDQCGTSDDSYSVGESGAG 128

Qy         135 SPVTGPGSKRPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTID 194
               :  |     ::  |||||||||||||||||||||||||||||||||||||:||||||||
Db         129 AGAAGPEMAKPMRPCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTID 188

Qy         195 KNRRKSCQACRLRKCYEVGMVKGGSRKDRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANL 254
               :|||||||||||||||||||:| | |::|   |::|| :
Db         189 RNRRKSCQACRLRKCYEVGMMKGGVRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANL 248

Qy         255 SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP--SAPPTEA 312
               | : | :  |   || :  :       :: |:|::| |  |||| :| ||   : |:||
Db         249 PQDRRKHISSSAGGGGGKSSVI      SMPPDQVLTLLPGAEPP-MLCSRQFVNRPYTF 306

Qy         313 ASMMMSITKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHP 372
               ::|  || :|||||||||:||||:|||:||| |||:|||| |:||||:||:|||  |
Db         307 VIVMTLLTSMADKELVHMIAWAKKLPGFLQLSLHDQVQLLESSWLEVLMIGLIWRSIEHP 366

Qy         373 GKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMY 432
               |||||| ||:||| || |||:| |||||||||| |||| ||| :|:::|:||:|||  :
Db         367 GKLIFAQDLILDRSEGDCVEGMAEIFDMLIATASRFRMLKLKPEEFVCLKAILLLNSGAF 426

Qy         433 PLVTATQD-ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMILSHVRHASN 491
               | | :    |  : :::: :||||:  | :|| |:|||| | | ||:|||||:|| ||
Db         427 SFCTGTMPPLHDSAAVQNMLDTITDALIHWINQSGCSAQQQSRRQAQLLLLLSHIRHMSN 486

Qy         492 KGMEHLTNMKCKNVVPVYDLLLEMLNAH-VLRGCPSSITGSEGS-----PAPQSPSPEQ 545
               |||||| :|||||  ||:|||||||||:|| | |  : : | |:         : :| |  |
Db         487 KGMEHLYSMKCKNFVPLYDLLLEMLDAHRVHRPDRPAEIWSQAIRELPEISRNSSSSSG 546
```

```
RESULT 33
ESR1_ORYLA
ID   ESR1_ORYLA     STANDARD;      PRT;   620 AA.
AC   P50241;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR OR NR3A1 OR MER.
OS   Oryzias latipes (Medaka fish) (Japanese ricefish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D-RR; TISSUE=Liver;
RA   Okada N., Kawahara T., Yamashita I.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kawahara I., Yamashita I.;
RT   "Oryzias latipes genomic DNA for estrogen receptor.",
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D28954; BAA25900.1; -.
DR   EMBL; AB033491; BAA86925.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    185       MODULATING.
FT   DNA_BIND    186    251       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     186    206       C4 TYPE.
FT   ZN_FING     222    246       C4-TYPE.
FT   DOMAIN      252    314       HINGE.
FT   DOMAIN      315    620       STEROID-BINDING.
FT   DOMAIN      299    307       POLY-GLY.
FT   DOMAIN      320    323       POLY-LEU.
FT   DOMAIN      511    515       POLY-LEU.
FT   DOMAIN      576    579       POLY-GLY.
SQ   SEQUENCE   620 AA;  67729 MW;  E9CBF190C2B28A522 CRC64;
```

```
  Query Match             42.7%;  Score 1199;  DB 1;  Length 620;
  Best Local Similarity   48.1%;  Pred. No. 9.1e-89;
  Matches  265;  Conservative   81;  Mismatches  129;  Indels   76;  Gaps   16;

Qy          36 YVDSHHEY----PAMI---------FYSPAVMNYSIPS--NVTNLEGGPGRQITSPNVLW 80
               :::  ::|    || |          :|| :   ||  :: :|  ||     ||| |
Db          62 FLEGTYDYAAPNPATTPLYSQSSTGYYSAPLETNGPPSEGSLQSLGSGP----TSPLVFV 117

Qy          81 PTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNR---CASPV 137
               |:   ||| :     || | |  :|   :          ::   :    |:|   |:
Db         118 PSSPRLSPFM--HPPSSHHYLETTSIPVYPSS--------HQGASPFDQGSPFDIGSLGF 167

Qy         138 TGPGS  -  --KRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQ 190
                | |:        :|  |||||||||||||||||||||||||||||||||||:||||||
Db         168 LGAGAGAGGPEMAKDTRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQ 227

Qy         191 CTIDKNRRKSCQACRLRKCYEVGMVKGGSRKDRGGRMLKHKRQRDDGEGRGEVGSAGD---- ----SAD 246
               ||||:|||||||||||||||:| | |::|   |::|| :
Db         228 CTIDRNRRKSCQACRLRKCYEVGMMKGGVRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRA 285

Qy         247 FQLHCAGKAKRS-GGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP--SAPPTEA 302
               : :|  |: : :|  ||        | : :: |||::| |  |||| :| ||   | |:||
Db         286 KIVHYICRKRSSIGGGGGGGGKISVISIPPFQVLTLQGAEPP-LLCSRQKISRPYTFV 344

Qy         303 SMMMSITKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG 362
               :||  || :|||||||||:||||:|||:||| ||| |||| |:||||:||:|||  ||
Db         345 TMMTLLTSMADKELVHMIAWAKKLPGFLQLSLHDQVLLLESSWLEVLMIGLIWRSIEHPG 404

Qy         363 KLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYP 422
               ||||| ||:|||:|| |||| |||||||| |||| ||| :|:::|:||:|||  :
```

```
Db          405 KLIFAQDLILDRNEGDCVEGMLELFDMLLATASRFRVLKLKPEEFVCLKAIILLNSGAFS 464

Qy          413 LVTATQDEADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNK 471
                |||:  :|   :|  :|:|||||:  |::||  :|:|: | | ||:||||:|| |||
Db          465 FCTGTMEPLHNSAAVQSMLDTITDALIHYISQSGYLAQFQARHQAQLLLLLSHIRHMSNK 524

Qy          472 GMEHLLNMKCKNVVPVYDLLLEMLNAHVL-  -RGCKS      ---- -SITGSEDSPAE 516
                ||||| :|||| ||:|||||||||:|| |    |  :|          | |   :|
Db          525 GMEHLYSMKCKNKVPLYDLLLEMLDAHRLHAPVSRAPQSLSQVDRDPPSTSSGGGQTAPGS 584

Qy          517 DSKSKEGSQNP 527
                | |:   ::|
Db          585 ISASRGRIESP 595


RESULT 34
ESR1_PAGMA
ID   ESR1_PAGMA      STANDARD;       PRT;   581 AA.
AC   O42182;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Pagrus major (Red sea bream) (Chrysophrys major).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
OC   Sparidae; Pagrus.
OX   NCBI_TaxID=143350;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Touhata K., Kinoshita M., Toyohara H., Sakaguchi M.;
RT   "Sequence and expression of a cDNA encoding the red seabream estrogen
RT   receptor.";
RL   Fisheries Sci. 64:131-135(1998).
CC    !  FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC    !  SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER
CC       BETA (BY SIMILARITY).
CC    !  SUBCELLULAR LOCATION: Nuclear.
CC    !  DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC    !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   -!- ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; AB007453; BAA22517.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   Prodom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc finger; Steroid-binding.
FT   DOMAIN          1     144       MODULATING (BY SIMILARITY).
FT   DNA_BIND      145     210       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING       145     165       C4-TYPE.
FT   ZN_FING       181     205       C4-TYPE.
FT   DOMAIN        211     272       HINGE.
FT   DOMAIN        273     581       STEROID-BINDING (BY SIMILARITY).
FT   DOMAIN        261     265       POLY-GLY.
FT   DOMAIN        278     281       POLY-LEU.
FT   DOMAIN        469     473       POLY-LEU.
FT   DOMAIN        537     545       POLY-GLY.
SQ   SEQUENCE   581 AA;  63461 MW;  D499BB6E7B4EB964 CRC64;

  Query Match             42.6%;  Score 1194.5;  DB 1;  Length 581;
  Best Local Similarity   47.5%;  Pred. No. 1.9e-88;
  Matches  257;  Conservative   88;  Mismatches  123;  Indels   73;  Gaps   15;

Qy           36 YVDSHHEYPAMT----------FYSPAVMNYSIPS-NVTNLEGGPGRQTTSPNVLW 80
                ::: ::| | |              :|| :  :  || :: :|  ||     || |
Db           17 FLEGTYDYAAPTPAPTPLYSHSTPGYYSAPLDAHGPPSDGSLQSLGSGP----NSPLVFV 72

Qy           81 PTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRK----VSGNRGASP 136
                |:  ||| :     :| | |  :|         ::| ::::: |:     | :  :
Db           73 PSSPRLSPFM- RPPTHHYLETTSIP- - VYRSSVPSSQQSVSREDQCGISDDSYSVG 125

Qy          137 VIGPGS-------KRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPAIN 189
                 :| |:        ::  |||||||||||||||||||||||||||||||||||:|||||
Db          126 ESGAGALAAGFEIAKEMREGACVSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPAIN 185

Qy          190 QCTIDKNRRKSCQACRLRKCYEVGMVKGGSKRERGGYRLVRRQRS -- -  - -- --- 234
                |||||:|||||||||||||||||:| | |::|   |::|| :
Db          186 QCTIDRNRRKSCQACRLRKCYEVGMMKGGMRKDR--GRVLRDDKQRPGLTSDPDKASKGLE 243

Qy          235 ------ADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP--S 286
                     | : | :  |   || :  :       :: |:|::| |  |||| :| ||   :
Db          244 HRTAPPQDRRKHISSSAGGGGGKSSMI------SMPPDQVLLLQGAEPP-MLCSRQKLN 296

Qy          287 APFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWR 346
                 |:|| :||  || :|||||||||:||||:|||:::||| |||:|||| |:|||||:||
Db          297 RPYIEVTMMTLLTSMADKELVHMIAWAKKLPGFLQLSLHDQVQLLESSWLEVLMIGLIWR 356

Qy          347 SIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILL 406
                || |||||||| ||:||| || |||| |||||||||| |||| ||| :|:::|:||:||
Db          357 SIBCPGKLIFAQDLILDRSEGDCVEGMAEIFDMLLATASRFRMLKLKPEEFVCLKAIILL 416

Qy          407 NSSMYPLVTATQD-ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHV 465
                || :  | | :      : :::: :|||:  | :|| |:|||| | | ||:|||:
Db          417 NSGAFSFCTGTMEPLHDGAAVQNMLDTITDALIHBINQSGCSAQQQSRRQAQLLLLLSHI 476

Qy          466 EHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH-VLRGCKSSITGSEC -SPAEDSKSEE 522
                || |||||||| :|||| ||:|||||||||:|| : |  : : | |:    |  |::
Db          477 RHMSNKGMEHLYSMKCKNKVPLYDLLLEMLDAHRIHRADRPAETWSQADREPPFTSRNSS 536

Qy          523 G 523
                |
Db          537 G 537


RESULT 35
ESR1_OREAU
ID   ESR1_OREAU      STANDARD;       PRT;   583 AA.
AC   P50240;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Oreochromis aureus (Israeli tilapia).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidei;
OC   Cichlidae; Oreochromis.
OX   NCBI_TaxID=47969;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97120581; PubMed=8961252;
RA   Tan N.S., Lam T.J., Ding J.L.;
RT   "Transcription regulatory signals in the 5' and 3' regions of
RT   Oreochromis aureus ER gene.";
RL   Mol. Cell. Endocrinol. 123:149-161(1996).
RN   [2]
RP   SEQUENCE OF 270-583 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=96272302; PubMed=8777315;
RA   Tan N.S., Lam T.J., Ding J.L.;
RT   "Molecular cloning and sequencing of the hormone-binding domain of
RT   Oreochromis aureus estrogen receptor gene.";
RL   DNA Seq. 5:359-370(1995).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X93557; CAA63774.1; -.
DR   EMBL; X93558; CAA63774.1; JOINED.
DR   EMBL; X93559; CAA63774.1; JOINED.
DR   EMBL; X93560; CAA63774.1; JOINED.
DR   EMBL; Z46665; CAA63774.1; JOINED.
DR   EMBL; Z46666; CAA63774.1; JOINED.
DR   EMBL; Z46667; CAA63774.1; JOINED.
DR   EMBL; Z46668; CAA63774.1; JOINED.
DR   EMBL; Z46669; CAA63774.1; JOINED.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER
DR   ProDom; PD000045; Znf_C4steroid; 1
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN          1    138       MODULATING (BY SIMILARITY).
FT   DNA_BIND      139    204       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING       139    159       C4-TYPE.
FT   ZN_FING       175    199       C4-TYPE.
FT   DOMAIN        205    265       HINGE.
FT   DOMAIN        266    583       STEROID-BINDING (BY SIMILARITY).
FT   DOMAIN        271    274       POLY-LEU.
FT   DOMAIN        461    465       POLY-LEU.
FT   DOMAIN        531    534       POLY-GLY.
SQ   SEQUENCE   583 AA;  64030 MW;  20CFE85CAFB36A0E CRC64;

  Query Match             42.2%;  Score 1184.5;  DB 1;  Length 583;
  Best Local Similarity   49.4%;  Pred. No. 1.2e-87;
  Matches  273;  Conservative   65;  Mismatches  128;  Indels   87;  Gaps   19;

Qy          6 SPSSLNSPSSYNCSQSIL----PLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPSN 61
              :|: | | |:  |  : |    || |||:
Db         40 APTPLYSHSTTGCYSAPLDAHGPLSDGSL-------------------------   58

Qy         62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNR 121
                :|  ||    ||| | |:   ||| :     || | |   :|   :    | || |
Db         59 -QSLGSGP----TSPLVFVPSSPRLSPFM--HPPSHHYLETTSTPVYRS---SHQ-PVPR 107

Qy        122 E----TLKRKVSCNRCASPVIGPGSKRPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSI 177
              |    |      |    :  |    ::  ||||||||||||||||||||||||||||||
Db        108 EDQCGIKDEAYSVGFLGAGAGGFEFTIKNIRFCAVCSDYASGYHYGVWSCEGCKAFFKRSI 167

Qy        178 QGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRK------ 231
              |||||||:|||||||||||||||||||||||||||||:| | |::|   |::||
Db        168 QGNNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKCGMRKERCGPVLRREKHGPA 225

Qy        232 QPSADEQL-HCAGF-AFPS          GGHAPEVREFLITDALSPPQLIVTILEAPP 277
              ||    : |  | |      ||:          ||  :       |: : |:|::| |  |||
Db        226 QPQTSQNLPTHKASPQDGRKPAMSSSSTSGPGGRSS------LNNMPPDQVITILQGAPP 279

Qy        278 PHVLISPP--SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
              | :| ||   | |:|| ::|  || :|||||||||:||||:|||::||| ||| |||| |
Db        280 PGLQSPQKMSRPYIEVIIMILLISMADKELVHMIIWAKKLPGFLQLSLHDQVILLESSW 338

Qy        336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHK 395
              :||||:||:|||| |||||| ||:|||:|| |||| ||||||||| ||| ||| : :
Db        339 LEVLMIGLIWRSIQCPGKLIFEEDLILDRNEGTCVEGMAEIFDMLLATVSRFRVLKLKPE 397

Qy        396 EYICVKAMILLNSSMYPLVTATQD-ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMR 454
              |:::|:||:|||||  :  | | :   |  : |:|: :|||||:: |:  | |:|||| |
Db        398 EEVCLKAILLLNSGAFSVTGTMEPLHDSVAVQHMLDLITDALIFHISHPGCSAQQQSRR 457

Qy        455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT---GS 510
              | ||:|||||:|| ||||||||  :||||| ||:|||||||||:|| : |  | | :   |
Db        458 QAQLLLLLSHIRHMSNKGMEHLYSMKCKNKVPLYDLLLEMLDAHRIHRPVKPSQSWSQGD 517

Qy        511 ECSPAEDSKSKEG 523
              ||   | |  |
Db        518 RDSPTASSTSSRG 530


RESULT 36
ESR1_MICUN
ID   ESR1_MICUN     STANDARD;      PRT:   525 AA.
AC   P57753;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha) (Fragment).
GN   ESR1 OR NR3A1 OR ESR.
OS   Micropogonias undulatus (Atlantic croaker).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
OC   Sciaenidae; Micropogonias.
OX   NCBI_TaxID=29154;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20461442; PubMed=11005855;
RA   Hawkins M.B., Thornton J.W., Crews D., Skipper J.K., Dotte A.,
RA   Thomas P.;
RT   "Identification of a third distinct estrogen receptor and
RT   reclassification of estrogen receptors in teleosts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:10751-10756(2000).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC    !  TISSUE SPECIFICITY. ABUNDANT IN THE LIVER, LESS ABUNDANT IN THE
CC       TESTES AND BARELY DETECTABLE IN THE OVARY AND BRAIN.
```

```
CC     !  DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC        A DNA BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC     !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC        NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF298183; AAG16713.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Steroid_hormn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; PARTIAL.
KW   Receptor; Transcription regulation; DNA binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   NON_TER       1      1
FT   DOMAIN       -1     82       MODULATING.
FT   DNA_BIND      83    148       NUCLEAR RECEPTOR TYPE.
FT   ZN_FING       83    103       C4 TYPE.
FT   ZN_FING      119    143       C4 TYPE.
FT   DOMAIN       149    209       HINGE.
FT   DOMAIN       210    525       STEROID BINDING.
FT   DOMAIN       192    197       POLY SER.
SQ   SEQUENCE   525 AA;  57637 MW;  2DD254E508370707 CRC64;

  Query Match             42.0%;  Score 1177;  DB 1;  Length 525;
  Best Local Similarity   54.4%;  Pred. No. 4.3e-87;
  Matches  247;  Conservative   61;  Mismatches  107;  Indels   40;  Gaps   10;

Qy         74 TSPNVLWPTPGHLSPLVVHRQLSHLYAEPQESPWCEARSLEHLLPVNRE----ILKREVS 129
              ||| |  |:   ||| :     || | |   :|   :         :::|    |     |
Db          2 TSPLVFVPSSPRLSPFM  HPPSHHYLETTSTPVYRSSVSSSQQQLSREDQCGTSDDSYS 59

Qy        130 -GNRCASPVTGPGS---KRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYIC 185
               |   |    | |     ::  |||||||||||| ||||||||||||||||||||||||:|
Db         60 VGESGAGAGAGAGGFFMAKEMRFCAVCSDYASGYDYGVWSCEGCKAFFKRSIQGHNDYMC 119

Qy        186 PATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR----------- 233
              |||||||||:|||||||||||||||:|||:| | |::|   |::|| :
Db        120 PATNQCTIDRNRRKSCQACRLRKCYQVGMMKGGVRKDR--GRVLRRDKRRTGISPKASKD 177

Qy        234 -----SADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
                  | : | :  :   ||   :       :|||:::| |  |||| || ||
Db        178 LEHRTAPPQDRRKHSSSSSSAGGGKSSI--  - IGMSPDQVILLLQGAEPP-VIYSRQK 231

Qy        286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLM 344
              | |:|| :||  || :|||||||||:||||:|||::||| |||:|||| |:||||:|||
Db        232 LSRPYTEVTMMTLLTSMADKELVHMIAWAKKLPGFIQLSLHDQVQLLESSWLEVLMIGLI 291

Qy        345 WRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMI 404
              ||||  |||||||| ||:||| || |||:|||||||||||||| |||: :|:::|:|:|
Db        292 WRSIHCPGKLIFAQDLILDRSEGDCVEGMAEIFDMLLATTSRFRMLKLKTEEFVCLKAII 351

Qy        405 LLNSSMYPLVTATQD ADSSPKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLS 463
              ||||  :  | | :       : ::|: :||||:  |::|| |:||||  | |:|:|||
Db        352 LLNSGAFSFCTGTMEPLHDGAAVQNMLDTITDALIRHISQSGCSAQQQSRRQAHLLLLLS 411

Qy        464 HVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
              |:|| |||||||| :||||| ||:|||||||||:||
Db        412 HIRHMSNKGMEHLYSMKCKNKVPLYDLLLEMLDAH 446

RESULT 47
ESR1_ONCMY
ID   ESR1_ONCMY      STANDARD;      PRT;   622 AA.
AC   P16058; Q9PSZ9; Q9PSZ8; Q9YGE6;
DT   01-APR-1990 (Rel. 14, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RX   MEDLINE 91006824; PubMed-2210031;
RA   Pakdel F., le Gac F., le Goff P., Valotaire Y.;
RT   "Full length sequence and in vitro expression of rainbow trout
RT   estrogen receptor cDNA.";
RL   Mol. Cell. Endocrinol. 71:195-204(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC   TISSUE=Ovary;
RX   MEDLINE 20114360; PubMed-10650938;
RA   Pakdel F., Metivier R., Flouriot G., Valotaire Y.;
RT   "Two estrogen receptor (ER) isoforms with different estrogen
RT   dependencies are generated from the trout ER gene.";
RL   Endocrinology 141:571-580(2000).
RN   [3]
RP   SEQUENCE OF 150-574 FROM N.A. (SHORT ISOFORM).
RC   TISSUE=Liver;
RX   MEDLINE 89127284; PubMed=2915648;
RA   Pakdel F., le Guellec C., Vaillant C., Le Roux M.G., Valotaire Y.;
RT   "Identification and estrogen induction of two estrogen receptors (ER)
RT   messenger ribonucleic acids in the rainbow trout liver: sequence
RT   homology with other ERs.";
RL   Mol. Endocrinol. 3:44-51(1989).
RN   [4]
RP   SEQUENCE OF 270-622 FROM N.A.
RX   MEDLINE 20159854; PubMed=10696781;
RA   Matthews J.B., Zacharewski T.R.;
RT   "Differential binding affinities PCBs, Bo-PCBs and Aroclors with
RT   recombinant human, rainbow trout (Onchorhynkiss mykiss) and green
RT   anole (Anolis carolinensis) estrogen receptors using a semi-high
RT   throughput competitive binding assay.";
RL   Toxicol. Sci. 53:326-339(2000).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/ER(L) (SHOWN HERE)
CC       AND A SHORT FORM/ER(S); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC    !  DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- MISCELLANEOUS: IN THE ABSENCE OF LIGAND, STEROID HORMONE RECEPTORS
CC       ARE THOUGHT TO BE WEAKLY ASSOCIATED WITH NUCLEAR COMPONENTS;
CC       HORMONE BINDING GREATLY INCREASES RECEPTOR AFFINITY. THE
CC       HORMONE-RECEPTOR COMPLEX APPEARS TO RECOGNIZE DISCRETE DNA
CC       SEQUENCES UPSTREAM OF TRANSCRIPTIONAL START SITES.
CC    !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M31559; AAA49552.1; -.
DR   EMBL; AJ242740; CAB45139.1; -.
DR   EMBL; AJ242741; CAB45140.1; -.
DR   EMBL; AF099079; AAD13610.2; -.
DR   PIR; A37197; A37197.
DR   PIR; A40070; A40070.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding; Alternative splicing.
FT   DOMAIN        1    192       MODULATING.
FT   DNA_BIND    193    258       NUCLEAR RECEPTOR-TYPE
FT   ZN_FING     193    213       C4-TYPE.
FT   ZN_FING     229    253       C4-TYPE.
FT   DOMAIN      259    323       HINGE.
FT   DOMAIN      324    622       STEROID-BINDING (POTENTIAL).
FT   DOMAIN      580    586       POLY-THR.
FT   VARSPLIC      1     45       MISSING (IN SHORT ISOFORM).
FT   CONFLICT     60     91       FNYLDGGYDYTAPAQGPAPLYYSTTPQDAHGP -> LTIWM
FT                                EGMTTQPLPKARPLSITPPPPRMPTD (IN REF. 1).
FT   CONFLICT    160    160       S -> A (IN REF. 1 AND 3).
FT   CONFLICT    289    298       YDIFHETAP -> TVTWSTQQF (IN REF. 1 AND
FT                                3).
FT   CONFLICT    335    337       AEP -> QT (IN REF. 1 AND 3).
FT   CONFLICT    449    449       P -> G (IN REF. 1 AND 3).
FT   CONFLICT    469    469       S -> P (IN REF. 1 AND 3).
FT   CONFLICT    564    564       S -> A (IN REF. 1 AND 3).
SQ   SEQUENCE   622 AA;  68084 MW;  AD70DD7136B5A6F3 CRC64;

  Query Match           41.5%;  Score 1165;  DB 1;  Length 622;
  Best Local Similarity   45.2%;  Pred. No. 5e-86;
  Matches  261;  Conservative   85;  Mismatches  150;  Indels   82;  Gaps   15;

Qy          4 KNSPSSLNSPSSYNCSQSILPLE--HGSIYIPSSYVDSHHEY------PAMTFYSPAVMN 55
              | ||    :|:|    :: | |   |      :|:|  ::|      || :|| :
Db         33 KLSPQQPTTPNS-----NMYPEETRGGGGAAAFNYLDGGYDYTAPAQGPAPLYYSTTPQD 87

Qy         56 YSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLV---VHRQLSH------LYAEPQKSP 106
                | :  :::   |   | | |   :   ||| :   :|    |        | |   :|
Db         88 AHGPPSDGSMQ-SLGSSPTGPLVFVSSSPQLSPQLSPFLHPPSHHGLPSQSYYLETSSTP 146

Qy        107 WCEARSLEHTLPVNRETL  - KRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYG 162
                :  : : |  : | |     |: | :   | |       :  :|||||:|||||||
Db        147 LYRSSVVTNQLSASEEKLCIASDRQQSYSAAGSGVRVEMANETRYCAVCSDFASGYHYG 206

Qy        163 VWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGLRKD 222
              ||||||||||||||||||||:||||||||:|:||||||||||||||||||| | |::
Db        207 VWSCEGCKAFFKRSIQGHNDYMCPATNQCTMDKNRRKSCQACRLRKCYEVGMVKGGLRKD 266

Qy        223 RGGYRLVPRRQ  -  ----        -  -  -    RSADEQLRCACRKAFRSCHATR 273
              | | |::|:                                |::  |: |    |  ||
Db        267 R-GGRVLKKDKRYCGPAGDRFKPYGDLEHRTAPPQDGGRNSSSSLNGQG-----GWRGPR 320

Qy        255 VRELLLDALSPEQLVLTLLEAEPPHVLISRP--SAPFTEASMMMSLTKLADKELVHMISW 312
              :       : |||::  |  ||||  | ||   : |:|| :||  || :|||||||||:|
Db        321 I------TMPPEQVIFLLQGAPPP-ALCSPQKVARPYTEVTMMTLLTSMADKELVHMIAW 373

Qy        313 AKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEG 372
              |||:|||| |||| |||:|||| |:|||||:||:|||| |||||||  ||:||| || ||||
```

```
Db        374 AKKVPGFQELSLHDQVQLLESSWLEVLMIGLIWRSIDHPGKLIFAQDLILDRSEGKCVEG 433

Qy        373 ILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVT-ATQDADSSRKLAHLLN 431
              : |||||||| |||| ||| :|:|:|:||:|||| :   : : :  :| :  :|:
Db        434 MAEIFDMLLAIVSRFRMLKLKPEEFVCLKAILLLNSGAFSFCSNSVESLHNSSAVESMLD 493

Qy        432 AVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLL 491
              :|||:  |: || | |||  | | ||:||||:|| |||||||| ::|||| ||:|||
Db        494 NITDALIHHISHSGASVQQQPRRQAQLLLLLSHIRHMSNKGMEHLYSIKCKNKVPLYDLL 553

Qy        492 LEMLNAHVLRGCKSSITGSEGSPAEDSKSEGSQNPQS 529
              ||||: | |:           || : ::: | :: | :
Db        554 LEMLDGHRLQ----------SPGKVAQAGEQTEGPST 580


RESULT 38
ESR1_ICTPU
ID   ESR1_ICTPU     STANDARD;      PRT;   617 AA.
AC   Q9YHZ7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha).
GN   ESR1 OR NR3A1 OR ESR.
OS   Ictalurus punctatus (Channel catfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC   Ictaluridae; Ictalurus.
OX   NCBI_TaxID=7998;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Ovary, and Liver;
RA   Patino R., Xia Z., Gale W.L., Maule A.G., Chang X.;
RT   "Novel transcripts of the estrogen receptor alpha gene in channel
RT   catfish.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=20218982; PubMed=10753576;
RA   Xia Z., Gale W.L., Chang X., Langenau D., Patino R., Maule A.G.,
RA   Densmore L.D.;
RT   "Phylogenetic sequence analysis, recombinant expression, and tissue
RT   distribution of a channel catfish estrogen receptor beta.";
RL   Gen. Comp. Endocrinol. 118:139-149(2000).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: OVARY AND TESTIS.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY
CC       NR3 SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF253505; AAG24543.1; -.
DR   EMBL; AF061275; AAC69548.1; -.
DR   EMBL; AF253506; AAG24544.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
```

```
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   PRINTS; PR00398; STRDBORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA binding; Nuclear protein;
KW   Zinc finger; Steroid binding; Alternative splicing.
FT   DOMAIN        1    186       MODULATING.
FT   DNA_BIND    187    252       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     187    207       C4 TYPE.
FT   ZN_FING     223    247       C4 TYPE.
FT   DOMAIN      253    315       HINGE.
FT   DOMAIN      316    617       STEROID-BINDING (POTENTIAL).
FT   DOMAIN      321    325       POLY-LEU.
FT   DOMAIN      512    516       POLY LEU.
FT   DOMAIN      577    591       POLY THR.
FT   VARSPLIC      1     36       MISSING (IN ISOFORM 2).
FT   VARSPLIC      1    425       MISSING (IN ISOFORM 3).
SQ   SEQUENCE   617 AA;  67670 MW;  7B208558E5EAD403 CRC64;

  Query Match             41.4%;  Score 1160.5;  DB 1;  Length 617;
  Best Local Similarity   49.5%;  Pred. No. 1.1e-85;
  Matches  259;  Conservative   77;  Mismatches  124;  Indels   63;  Gaps   16;

Qy         28 GSIYIPSSYVDSHHEY---PAMTFYSPAVMNYSI---PSNVTNLEGGPGRQTTSPNVL 79
              |:|    |:|     |   |  |  : :|  ||:       |:  |  |    ::|| |
Db         47 GGISSIAHYLDGTENYTTNPDAT--NSSVDYYSVAPEPQEENLQPLPNG----SSSPPVF 100

Qy         80 WPTPGHLSPLVVH-----RQLSHLYAEP-----QKSPWCEARS-LEPTLPVNRETLK 125
                |:   ||| : |       |    | ||      : |    ( ( :|      |: : :
Db        101 VPSSPQLSPFLGHPPAGQHTAQQVPYYLEPSGTSIYRSSVLASAGSRVELCSAPGRQDVY 160

Qy        126 RKV--SGNRCASPVTGP-GSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHND 182
              |  ||   ||  :|  |  ::  :|:||||||||||||||||||||||||||||||||
Db        161 TAVGASGPSGASGPSGAIGLVKETRYCSVCSDYASGYHYGVWSCEGCKAFFKRSIQGHND 220

Qy        183 YICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSPREPCGYPLVPRQPSADEQLHCA 242
              |:||||||||||:|||||||||||||||||||||:| | |:|| | |::  |        :
Db        221 YVCPATNQCTIDPNRRKSCQACRLPKCYEVGMMKGGPRKER-GGRLIKHNRRP- ----S 273

Qy        243 GKAKRSGGHAPR---VRELL-     -  ------ --LDALSPEQLVLTLLEAEPPHVL 281
              |  :|  |::      |||  |              :  :||||::| || ||||  |
Db        274 GLEEPERGYSKAQSGSDVRFALPQDGQSSSCIGCCVADVVCMSPEQVLLLLLRAEPP TL 332

Qy        282 ISRP--SAPETEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVL 339
                 ||   | |::| ::|  || :||:|||||:||||:||| :||| |||:|||| |:|:|
Db        333 CSRQKHSRPYSELTIMSLLTNMADRELVHMIAWAKKVPGFQDLSLHDQVQLLESSWLEIL 392

Qy        340 MMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLC 399
              |:||:|||  |||||||  ||:|| |||:|||: |||||||||| :||| |||:|||| |:|:|
Db        393 MIGLIWRSIYHPGKLIFAQDLILDKSEGECVEGMAEIFDMLLATVARFRELKLKKEEFVC 452

Qy        400 VKAMILLNSSMYPLVTA-TQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL 458
              :||:|||||  :   ::  :        :  ::: :||||:: |:|||| | || | | |
Db        453 LKAIILLNSGAFSFCSSPVEPLRDGFMVQCMMDNITDALIYYISQSGISVQLQSRRQAQL 512

Qy        459 LMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLR 501
              |:|||:|| | |||||| :|||| ||:|||||||:|| ||
Db        513 LLLLSHIRHMSYKGMEHLYSMKCKNKVPLYDLLLEMLDAHRLR 555


RESULT 39
ESR1_SALSA
ID   ESR1_SALSA     STANDARD;      PRT;   595 AA.
AC   P50242;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha) (Fragment).
GN   ESR1 OR NR3A1 OR ESR.
OS   Salmo salar (Atlantic salmon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX   NCBI_TaxID=8030;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Rogers S.A., Llewellyn L., Ramsurn V.P., Sweeney G.E., Wigham T.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC    !- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89959; CAA61999.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   NON_TER       1      1
FT   DOMAIN       -1    104       MODULATING (POTENTIAL).
FT   DNA_BIND    105    170       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     105    125       C4-TYPE.
FT   ZN_FING     141    165       C4-TYPE.
FT   DOMAIN      171    246       HINGE.
FT   DOMAIN      237    545       STEROID-BINDING (POTENTIAL).
SQ   SEQUENCE   545 AA;  58900 MW;  669F13D7B243B612 CRC64;

  Query Match             40.4%;  Score 1134.5;  DB 1;  Length 545;
  Best Local Similarity   48.8%;  Pred. No. 1.2e-83;
  Matches  246;  Conservative   73;  Mismatches  114;  Indels   71;  Gaps   15;

Qy         70 GRQTTSPNVLWPTPGHLSPLVVHRQLSH------LYAEPQKSPWCEARSLEHTLPVNRET 123
              |   | | |   :  |||  :|    |       | |   :|   :  : : | : |
Db         17 GSSPTGPLVFVSSSPQLSPF-LHPPGHGLPSQSYYLETSSTPLYRSSVVTNQLSASEE- 74

Qy        124 LKRKVSGNRCAS-PVTGPGSK--    RDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSI 177
                |  :: ||  |    | | :       :  :|||||:|||||| ||||||||||||||
Db         75  KLCTTSNRQQSYAAAGSGVRVFEMANETRYCAVCSDFASGYHYGFWSCEGCKAFFKRSI 133

Qy        178 QGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSPREPCGYPLVPRQKSADE 237
              ||||||:|||||||:|:|||||||||||||||||||||| | |::| | |::|:     :
Db        134 QGHNDYMCPATNQCTMDRNRRKSCQACRLRKCYEVGMVKGGLRKDR-GGRVLRK --- D 187

Qy        238 QLHC--AGKAKR----------------------SGGH---APRVRELLLDALSPEQL 268
              : :|  ||  ::                        |||     ||:       : |||:
Db        188 KRYCGPAGDREKPYGDLEHRTAPPQDGGRNSSSSSLSGGGWCGPRI- ----TMPPEQV 241

Qy        269 VLTLLEAEPPHVLISRP--SAPETEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFD 326
```

```
Db        242 LFLLQCAEPP-AICSRQKVARPYTEVTMMTLLTSMADKELVHMIAWAKKVPGFQELSLHD 300

Qy        327 QVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSR 386
              ||:||||| |:||||:||:|||| ||||||| ||:||| || |||| ||||||||| ||
Db        301 QVQLLESSWLEVLMIGLIWRSIHCPGKLIFAQDLILDRSEGDCVEGMAEIFDMLLATVSR 360

Qy        387 FRFLKLQHKEYLCVKAMILLNSSMYPLVT-ATQDADSSRKLAHLLNAVTDALVWVIAKSG 445
              || |||: :|:::|:||:|||||  :   : : :   :|  :  :|: :||||:  |: ||
Db        361 FRMLKLKPEEFVCLKAIILLNSGAFSFCSNSVESLHNSSAVESMLDNITDALIHHISHSG 420

Qy        446 ISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKS 505
              | |||  |   ||:||||:|| |||||||| ::|||| ||:|||||||: | |:
Db        421 ASVQQQPRRQVQLLLLLSHIRHMSNKGMEHLYSIKCKNKVPLYDLLLEMLDGHRLQ---- 476

Qy        506 SITGSECSPAEDSKSKEGSQNPQS 529
                    || : ::: | :: | :
Db        477 -------SPGKVAQAGEQTEGPST 493


RESULT 40
ESR1_CNEUN
ID   ESR1_CNEUN     STANDARD;      PRT;   307 AA.
AC   Q91424;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Estrogen receptor (ER) (Estradiol receptor) (ER-alpha) (Fragment).
GN   ESR1 OR NR3A1 OR ESR
OS   Cnemidophorus uniparens (Whiptail lizard).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Teioidea;
OC   Teiidae; Cnemidophorus.
OX   NCBI_TaxID=37197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oviduct, and Kidney;
RX   MEDLINE=96103562; PubMed=7495707;
RA   Young L.J., Godwin J., Grammer M., Gahr M., Crews D.;
RT   "Reptilian sex steroid receptors: amplification, sequence and
RT   expression analysis.";
RL   J. Steroid Biochem. Mol. Biol. 55:261-269(1995).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBUNIT: BINDS DNA AS A HOMODIMER. CAN FORM A HETERODIMER WITH ER-
CC       BETA (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S79923; AAB35739.1; -.
DR   HSSP; P03372; 1HCP.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; PARTIAL.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   NON_TER       1      1
FT   DNA_BIND     <1     36       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING       7     31       C4-TYPE.
FT   DOMAIN       37     95       HINGE.
FT   DOMAIN       96    307       STEROID-BINDING.
SQ   SEQUENCE   307 AA;  35202 MW;  D80A1F2FE396AC61 CRC64;

  Query Match             41.3%;  Score 877;  DB 1;  Length 307;
  Best Local Similarity   54.8%;  Pred. No. 2.9e-63;
  Matches  172;  Conservative   62;  Mismatches   54;  Indels   26;  Gaps    7;

Qy        179 CHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGIRKDRRGGRILKHKRQREEH 238
              |||||:|||||||||||||||||||||||||||||:| | |::| | |::: :|  :|
Db          1 CHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRILKHKRQREEH 60

Qy        239 LHCAGKAKRSGG-----HAPRV--RELLLD---------ALSPEQLVLTLLEAEPPHVLI 282
                    |: |      :| :    |::           :|: :|:|  |||||||  |:
Db         61 ------DNRNAGAIVERRSPNLWPSPLMITHNKKNSPALSLTADQIVSALLEAEPP-VVY 113

Qy        283 SR--PSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLM 340
              |   || ||:|||||  || |||:|||||:|||::||||:||| ||| |||  |:|:||
Db        114 SEYDPSRPFSEASMMTLLTNLADRELVHMINWAKRVPGFVDLSLHDQVHLLECAWLEILM 173

Qy        341 MGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCV 400
              :||:|||::|||||:|||:|:|||::|||||| :|||||||||:|||| : :| :|:::|
Db        174 IGLVWPSVEHPGKLIFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNVQGEEFVCL 233

Qy        401 KAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLL 459
              |::||||| :|  :::| :  :    :  :|: : | |: ::|||:| |||  ||| ||
Db        234 KSIILINSGIYTFISSTLKSLEEKDHIHRVLDKITDTLIHLMAKSGLSLQQQHRRLAQLL 293

Qy        460 MLLSHVRHASNKGM 473
              ::||| || |||||
Db        294 LILSHFRHMSNKGM 307


Search completed: June 16, 2003, 13:13:36
Job time : 14.4083 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:34:03 ; Search time 30.5411 Seconds
(without alignments)
3575.671 Million cell updates/sec

Title: US-08-826-361A-25
Perfect score: 2805
Sequence: 1 MDIKNSPSSLNSPSSYNCSQ ECSPAPDSKSKEQSQNPQSQ 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	2420	86.3	499	6	Q95MF0	Q95mf0 macaca arct
2	2317	82.6	486	6	Q95ME9	Q95me9 callithrix
3	2273	81.0	503	11	Q91Z86	Q91z86 mus musculu
4	1691	60.3	542	13	Q90ZE6	Q90ze6 squalus aca
5	1454.5	51.9	553	13	Q90WS8	Q90ws8 brachydanio
6	1410.5	50.3	565	13	Q8QHK9	Q8qhk9 paralichthy
7	1397.5	49.8	592	13	Q90WS9	Q90ws9 brachydanio
8	1389.5	49.5	553	13	Q98SM7	Q98sm7 brachydanio
9	1366	48.7	562	13	Q8QW75	Q8qw75 oryzias lat
10	1243	44.3	250	6	Q95JC0	Q95jc0 canis famili
11	1235.5	44.0	581	13	Q8QWA9	Q8qwa9 cnemidophor
12	1216.5	43.4	587	13	Q8QWB0	Q8qwb0 caiman croc
13	1210	43.1	569	13	Q98SM9	Q98sm9 brachydanio
14	1188	42.4	564	13	Q90WV1	Q90wv1 carassius a
15	1166.5	41.6	578	13	Q8QHL0	Q8qhl0 paralichthy
16	1165.5	41.6	431	6	Q95L13	Q95l13 ovis aries
17	1157	41.2	620	13	Q90WB6	Q90wb6 clarias gar
18	1118	39.9	554	13	Q90ZM8	Q90zm8 petromyzon
19	1099	39.2	458	13	Q9DDJ3	Q9ddj3 halichoeres
20	1061.5	37.8	391	13	Q98SM8	Q98sm8 brachydanio
21	942.5	33.6	335	13	Q9DDZ4	Q9ddz4 micropterus
22	679	24.2	276	13	Q91984	Q91984 brachydanio
23	594.5	21.2	200	13	Q90Z59	Q90z59 brachydanio
24	544	19.4	139	11	Q9JM97	Q9jm97 mus musculu
25	527.5	18.8	484	5	Q9VSE9	Q9vse9 drosophila
26	527.5	18.8	496	5	Q8WS79	Q8ws79 drosophila
27	525	18.7	101	11	Q923V1	Q923v1 mesocricetu
28	517.5	18.4	159	6	O46402	O46402 sus scrofa
29	506	18.0	249	6	Q95JC1	Q95jc1 canis famil
30	504	18.0	710	13	Q91BD5	Q91bd5 anguilla ja
31	491	17.5	583	13	Q9DEV4	Q9dev4 xenopus lae
32	489	17.4	703	13	Q8UVY3	Q8uvy3 xenopus lae
33	475.5	17.0	323	4	Q96I02	Q96i02 homo sapien
34	468.5	16.7	732	13	Q9DDU9	Q9ddu9 xenopus lae
35	467.5	16.7	689	13	Q8QGX5	Q8qgx5 anguilla ja
36	466.5	16.6	438	13	Q90ZM7	Q90zm7 petromyzon
37	458.5	16.3	525	6	Q95L58	Q95l58 mustela vis
38	458	16.3	436	5	Q8T5C6	Q8t5c6 biomphalari
39	458	16.3	939	6	Q9GLW0	Q9glw0 canis famili
40	456.5	16.3	790	13	P70048	P70048 xenopus lae
41	452.5	16.1	414	11	Q8VCR0	Q8vcr0 mus musculu
42	452	16.1	848	13	Q9YGV9	Q9ygv9 anguilla ja
43	449.5	16.0	400	5	O61448	O61448 amblyomma a
44	448.5	16.0	688	13	Q8UWB8	Q8uwb8 oreochromis
45	446	15.9	336	4	Q14267	Q14267 homo sapien
46	445	15.9	982	6	Q9N0W8	Q9n0w8 saimiri sci
47	442	15.8	87	6	Q9N0K2	Q9n0k2 sus scrofa
48	442	15.8	155	13	Q9DDZ3	Q9ddz3 micropterus
49	441.5	15.7	401	13	Q90ZM6	Q90zm6 petromyzon
50	440.5	15.7	360	13	O42274	O42274 crocodylus
51	436	15.5	412	13	Q91B40	Q91b40 xenopus lae
52	436	15.5	414	5	O61449	O61449 amblyomma a
53	435	15.5	363	5	Q9UAF1	Q9uaf1 polyandroca
54	434.5	15.5	906	4	Q9UN21	Q9un21 homo sapien
55	434	15.5	895	6	Q9GKN9	Q9gkn9 sus scrofa
56	434	15.5	896	6	Q9GKL7	Q9gkl7 sus scrofa
57	433.5	15.5	793	11	Q8R5J0	Q8r5j0 rattus norv
58	433.5	15.5	794	11	Q8R463	Q8r463 rattus norv
59	433	15.4	87	6	Q8SPK8	Q8spk8 equus cabal
60	432.5	15.4	730	4	Q13771	Q13771 homo sapien
61	431.5	15.4	853	13	O93245	O93245 oncorhynchu
62	429.5	15.3	427	5	Q9NG48	Q9ng48 apis mellif
63	427.5	15.2	452	13	Q91613	Q91613 xenopus lae
64	427.5	15.2	692	13	Q9W6F4	Q9w6f4 haplochromi
65	426	15.2	763	13	Q8QFV7	Q8qfv7 haplochromi
66	425	15.2	340	5	Q95WF8	Q95wf8 acropora mi
67	425	15.2	854	13	O93244	O93244 oncorhynchu
68	424.5	15.1	438	13	Q90418	Q90418 brachydanio
69	423.5	15.1	334	4	Q9H2M1	Q9h2m1 homo sapien
70	423	15.1	475	5	O76241	O76241 aca pugilat
71	421.5	15.0	769	13	O93497	O93497 paqrus majo
72	421.5	15.0	839	13	Q8QFV2	Q8qfv2 carassius a
73	419	14.9	379	13	Q90415	Q90415 brachydanio
74	417	14.9	563	13	Q9DDJ4	Q9ddj4 halichoeres
75	416.5	14.8	427	13	Q8UUM6	Q8uum6 oryzias lat
76	415.5	14.8	448	5	Q95WQ3	Q95wq3 acropora mi
77	414	14.8	797	13	Q9PWG5	Q9pwg5 anguilla ja
78	413.5	14.7	422	13	Q90417	Q90417 brachydanio
79	411.5	14.7	303	6	O97684	O97684 ovis aries
80	409	14.6	389	5	Q9U7D9	Q9u7d9 locusta mig
81	409	14.6	600	13	Q98TQ3	Q98tq3 oryzias lat
82	408	14.5	380	5	Q95WG4	Q95wq4 acropora mi
83	404.5	14.4	344	13	Q91445	Q91445 serinus can
84	403	14.4	408	5	Q9NFY1	Q9nfy1 tenebrio mo
85	403	14.4	763	13	Q8UWB7	Q8uwb7 oreochromis
86	399	14.2	322	11	Q64116	Q64116 rattus norv
87	396	14.1	539	11	Q9QWM0	Q9qwm0 rattus norv
88	396	14.1	560	11	Q9QWM1	Q9qwm1 rattus norv
89	391	13.9	585	4	Q16311	Q16311 homo sapien

90	387.5	13.8	467	5	Q96215	Q96215 lucilia cup
91	385	13.7	541	4	Q9UEC0	Q9uec0 homo sapien
92	385	13.7	643	6	O97726	O97726 sus scrofa
93	383	13.7	448	13	Q92019	Q92019 xenopus lae
94	383	13.7	459	5	Q9QSG7	Q9qsq7 aedes aegyp
95	383	13.7	484	5	Q9QSG8	Q9qsq8 aedes aegyp
96	381	13.6	127	6	Q9BDX3	Q9bdx3 capra hircu
97	380	13.5	462	11	P97513	P97513 mus spretus
98	378.5	13.5	485	5	Q9U3Y3	Q9u3y3 aedes albop
99	378.5	13.5	627	11	Q9QZB6	Q9qzb6 mus musculu
100	377	13.4	495	6	Q9GKL1	Q9gkl1 equus caball
101	377	13.4	680	5	Q9U3Y4	Q9u3y4 aedes albop
102	376.5	13.4	443	11	Q8VHB6	Q8vhb6 mesocricetu
103	376	13.4	458	11	Q91VK5	Q91vk5 mus musculu
104	375	13.4	517	13	Q90YL6	Q90yl6 brachydanio
105	374.5	13.4	516	13	O42186	O42186 brachydanio
106	374	13.3	474	5	Q8ST32	Q8st32 dirofilaria
107	373	13.3	455	11	Q9JJI9	Q9jji9 tamias sibi
108	373	13.3	458	11	Q91YX2	Q91yx2 mus musculu
109	372	13.3	468	13	Q91545	Q91545 xenopus lae
110	371	13.2	448	13	Q91425	Q91425 cnemidophor
111	371	13.2	469	13	Q91B81	Q91b81 rana ruqosa
112	371	13.2	501	13	Q91B82	Q91b82 rana ruqosa
113	371	13.2	872	5	Q8SY10	Q8sy10 drosophila
114	370.5	13.2	426	13	Q90968	Q90968 gallus gall
115	370.5	13.2	459	11	Q9QWJ1	Q9qwj1 rattus norv
116	370	13.2	501	13	Q91544	Q91544 xenopus lae
117	368.5	13.1	435	5	O96562	O96562 tripedalia
118	368.5	13.1	731	5	Q9BPL0	Q9bpl0 schistosoma
119	367.5	13.1	444	13	Q91391	Q91391 brachydanio
120	367.5	13.1	444	13	Q90271	Q90271 brachydanio

ALIGNMENTS

```
RESULT 1
Q95MF0
ID   Q95MF0       PRELIMINARY;       PRT;   499 AA.
AC   Q95MF0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta 2.
GN   ERBETA2.
OS   Macaca arctoides (Stump-tailed macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9540;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Scobie G.A., Wilson J.A., Millar M.R., Macpherson S., Saunders P.T.;
RT   "The estrogen receptor beta variant ERBeta cx/ERBeta2 is expressed in
RT   a wide range of tissues in both Old and New World primates.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF393815; AAK71417.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Ptam; PF00104; hormone_rec; 1.
DR   Ptam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   499 AA;  55782 MW;  A91DA345C871BC0A CRC64;

  Query Match             86.3%;  Score 2420;  DB 6;  Length 499;
  Best Local Similarity   95.6%;  Pred. No. 6.3e-216;
  Matches  456;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVESHHEYPAMTFYSPAVMNYSIPS 60

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEAKSLEHTLPVN 120
              |||||||||||||| ||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 NVTNLEGGPGRQTASPNVLWPTPGHLSPLAVHRQLSHLYAEPQKSPWCEAKSLEHTLPVN 120

Qy        121 PETLKRKVSGNPCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
              |||||||||||||| |||| | ||||||||||||:||||||||||||||||||||||||||
Db        121 RETLKRKVSGNRCTSPVTSPSSKRDAHFCAVCSDFASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQVH 240

Qy        241 CAGKAKRSGCHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
              || |||||| | | ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CASKAKRSGSHTPLVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300

Qy        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDL 360

Qy        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
              |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSNMYPLVTATQDA 420

Qy        421 DSSKKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLL 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||  :  | |
Db        421 DSSKKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHAKAEKASHSL 477
```

```
RESULT 2
Q95ME9
ID   Q95ME9       PRELIMINARY;       PRT;   486 AA.
AC   Q95ME9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta 2.
GN   ERBETA2.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Scobie G.A., Wilson J.A., Millar M.R., Macpherson S., Saunders P.T.;
RT   "The estrogen receptor beta variant ERBeta cx/ERBeta2 is expressed in
RT   a wide range of tissues in both Old and New World primates.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF393816; AAK71418.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Ptam; PF00104; hormone_rec; 1.
DR   Ptam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   486 AA;  54344 MW;  4B29385B7471F277 CRC64;

  Query Match             82.6%;  Score 2317;  DB 6;  Length 486;
  Best Local Similarity   91.0%;  Pred. No. 2.2e-206;
  Matches  434;  Conservative   18;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              ||||||||||||||||  |||||||| ||||||||:|||||||||||||||||||||||||
Db          1 MDIKNSPSSLNSPSSYNFGQSILPLEHGPIYIPSSYVESHHEYPAMTFYSPAVMNYSIPS 60
```

```
Qy          61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPIVVHRQLSHIYAEPQKSPWCEARSLEHTLPVN 120
               :||||| ||||| |||:|:|| |||||||| || ||||||||||||||||||||||||||:
Db          61 SVTNLEEGPGRQTTSPNMIWSTPGHLSPLAVHRQLSHLYAEPQKSPWCEARSLEHTLPVS 120

Qy         121 EPTLKRKVSGNPGASPVTGPGSKRPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
               |||||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||
Db         121 EPTIKRKVSGNPGASPVTGPSSKRPDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180

Qy         181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRPQRSADFQLH 240
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :|:||||
Db         181 NDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRPQCNAFFQLH 240

Qy         241 CAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMMSLTK 300
               |||||||||| ||||||| |||||||||||||||||||||||| ||||||||||||||||
Db         241 CAGKAKRSGGHVPRVRELLLSALSPEQLVLTLLEAEPPHVLISRPSVPFTEASMMMSLTK 300

Qy         301 LADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDL 360
               |||:|||||||||||||||||| |||||||||:||||:|||||||||||||||||||:|
Db         301 LADPELVHMISWAKKIPGFVELSLLDQVRLLESCWLEVLMVGLMWRSIDHPGKLIFAPNL 360

Qy         361 VLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDA 420
               :|||||||||||||:|||||||||||||||||||||||||:|||   |||||||||
Db         361 LLDRDFGKCVEGILEVFDMLLATTSRFRELKLQHKEYICVKAMVLLNSQVYPLVTATQDA 420

Qy         421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLL 477
               :||:|||||||||||||||| ||||||:|||||||||:||| :  | |
Db         421 ESSQKLAHLLNAVTDALVWVIAKSGFSSQQQSVRLANLLMILSHIRHASNKGMEHLL 477


RESULT 3
Q91Z86
ID   Q91Z86      PRELIMINARY;      PRT;   503 AA.
AC   Q91Z86;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta type II splice variant.
GN   ESR2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DDY;
RA   Dang Z., van Bezooijen R., Karperien M., Papapoulos S., Lowik C.;
RT   "Exposure of KS483 cells to estrogen enhances osteogenesis and
RT   inhibits adipogenesis.";
RL   J. Bone Miner. Res. 0:0-0(2001).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AY054413; AAL15175.1; -.
DR   MGD; MGI:109392; Esr2.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   503 AA;  56072 MW;  D68548BFD3693251 CRC64;

  Query Match             81.0%;  Score 2273;  DB 11;  Length 503;
  Best Local Similarity   85.7%;  Pred. No. 2.8e-202;
  Matches  431;  Conservative   22;  Mismatches   32;  Indels   18;  Gaps    1;

Qy          46 MTPYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVIWPTPGHLSPIVVHRQLSHIYAEPQKS 105
               |||||||||||:||:  |||||| ||| |||||||| ||||||  | | | |||||||
Db           1 MTPYSPAVMNYSVPSSTGNLFGGPVRQTASPNVIWPTSGHLSPIATHCQSSLIYAEPQKS 60

Qy         106 PWCEARSLEHTLPVNRPTLKRKVSGNPGASPVTGPGSKRPDAHFCAVCSDYASGYHYGVWS 165
               |||||||||||||||||||: |: |||||| | :|||||||||||||||||||||||||

Db          61 PWCEARSLEHTLPVNREILKRKLGGSGGASPVTSPSAKRDAHFCAVCSDYASGYHYGVWS 120

Qy         166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 225
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCG 180

Qy         226 YRLVRPQRSADFQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
               |||:|||||| ||:||  |||||: || |||:|||||:|||||||||||||||||:|||
Db         181 YRIVRPQPSASFQVHCLNKAKRTSGHTPRVKELLLNSLSPEQLVLTLLEAEPPNVLVSRP 240

Qy         286 SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
               | ||||||||||||||||||| |||||||||| |||||||||||| ||||||||||||:|||
Db         241 SMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVELSLLDQVRLLESCWMEVLMVGLMW 300

Qy         346 RSIDHPGKLIFAPDLVLD----------------RDEGKCVEGILEIFDMLLATTSRF 387
               |||||||||||||||||                |||||||||||| |||||||||
Db         301 RSIDHPGKLIFAPDLVLDRSSEDPHWHVAQIKSAVPPDEGKCVEGILEIFDMLLATTARF 360

Qy         388 RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSPKLAHLLNAVTDALVWVIAKSGIS 447
               |||||||||||||||||||||||||| |||:|:|||||| |||||||||||||||||||||
Db         361 RELKLQHKEYLCVKAMILLNSSMYPLATASQEAESSPKLTHLLNAVTDALVWVISKSGIS 420

Qy         448 SQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSI 507
               |||||:|||||||||||| |||||||||:||||||||||||||||||| ||| ||||
Db         421 SQQQSVRLANLLMILSHVRHTSNKGMEHLLSMKCKNVVPVYDLLLEMLNAHILRGYKSSI 480

Qy         508 IGSECSPAEDSKSKEGSQNPQSQ 530
               :|||||   |||||||||||| |||
Db         481 SGSECCSTEDSKSKEGSQNLQSQ 503


RESULT 4
Q90ZE6
ID   Q90ZE6      PRELIMINARY;      PRT;   542 AA.
AC   Q90ZE6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta.
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Squaloidei; Squalidae; Squalus.
OX   NCBI_TaxID=7797;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sikora R., Betka M., Callard G.V.;
RT   "Molecular Cloning and Stage-Related Expression of an Estrogen
RT   Receptor Beta Subtype During Spermatogenesis in the Dogfish Shark
RT   (Squalus Acanthias) Testis.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF147746; AAK57823.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   542 AA;  61285 MW;  D539A29A9718DD37 CRC64;

  Query Match             60.3%;  Score 1691;  DB 14;  Length 542;
  Best Local Similarity   61.7%;  Pred. No. 3.2e-148;
  Matches  320;  Conservative   83;  Mismatches   98;  Indels   18;  Gaps    7;

Qy           3 IKNSPSSLNSPSSYNCSQSTLP---LEHQSTYIPSSYVDSHHEYPAMTEYSPAVMNYSIPS 59
               ||||| | :   |:   | ||   ||| : |||||:: ||:| : ||||::: ||:||
Db          27 IKNSPSQLTPQPQYS---STIPGLSEHGPVCIPSSYVENRHEFPTLAEYSPSTIGYSMPS 83

Qy          60 NVTNLEGGPGRQTTSPNVIWPTPGHLSPIVVHRQLSHIYAEPQKSPWCEARSLEHTLPVN 120
               : :  :|   ||: |||:: | : ||:||: :| |   :|||| |||| : || :  | :|
```

```
Db          84 DASGPDCTTVRQSLSPSMYWSSTGHVSPTTLDCQQPTMYAEPPKSPWDDLRSGDQQLFLN 142

Qy         121 RETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGH 180
               || ||:| :    | | |   |:||||||||||:|:|||||||||||||||||||||||||
Db         143 RENLKKKPAPP--GSSVNGVCSRRDAHFCAVCNDFASGYHYGVWSCEGCKAFFKRSIQGH 200

Qy         181 NDYICPATNQCTIDKNRKKSCQACRLRKCYEVGMVKCGSRKERCGYRLVKRQRSADEQLH 240
               | ||||||||||||||||||||||||||||||||:|  :||:|| ||| |: | :  |
Db         201 NAYICPATNQCTIDKNRKKSCQACRLRKCYEVGMMKSEIRRDRCSYRLLKQNRLGSSAQAQ 260

Qy         241 CAGKAKRSG GHAERVRELLLLDALSPEQLVLTLLEAEPPHVL-ISRPSAPFTEASMMMSL 298
                 : ||||      : :  |: |:||:|: :||||||||:  :: |: |:|| ||||||
Db         261 HWSRGKPSGESDFTNISLDCLNELAPEELLSSLLEAEPPNVYSLNHPNKPYTEVSMMMSL 320

Qy         299 TKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAP 358
               | |||:|||||:||||:|||||| | |||:||| ||:||||:|||||||::||||:|||
Db         321 TNLADRELVHMIAWAKKVPGFVELDLHDQVQLLECCWLEVLMVGLMWRSIEYPGKLLFAP 380

Qy         359 DLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMVPLVTATQ 418
               ||:|||||:|||||||||||||| ||||:|||||:||||:|||:|||||||:|    |:
Db         381 DLILDRDEGQCVEGILEIFDMLLAATSRFRDLKLQHEEYLCLKAMVLLNSSMFPRSGVTE 440

Qy         419 DADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLN 478
               : ::  ||  :|: :|| |:| ::|||| |||: |||:|||||||:|||||||||| :
Db         441 EHENREKLHKILDTTTDTLIWCMSKSGIPPQQQATRLAHLLMLLSHIRHASNKGMEHLYS 500

Qy         479 MKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAED 517
               |||||||| ||||||||:|||:          |   |::|
Db         501 MPCPNVVPFYDLLLEMLDAHVIY        -SPTPMSAD 538


RESULT 5
Q90WS8
ID   Q90WS8       PRELIMINARY;      PRT;   553 AA.
AC   Q90WS8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ZIER beta2 protein.
GN   ZFER BETA2.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Menuet A., Pellegrini E., Anglade I., Blaise O., Laudet V., Kah O.,
RA   Pakdel F.;
RT   "Binding characteristics, transactivation properties and central
RT   expression of three estrogen receptor subtypes in zebrafish.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AJ414567; CAC93849.1; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc finger.
SQ   SEQUENCE   553 AA;  62323 MW;  84C4DE6C7AC0C5ED CRC64;

  Query Match             51.9%;  Score 1454.5;  DB 14;  Length 553;
  Best Local Similarity   55.7%;  Pred. No. 4e-126;
  Matches  290;  Conservative   78;  Mismatches  126;  Indels   27;  Gaps    9;

Qy          16 YNCSQSILPLEHGSIYIPSSYVDSHHEYPAMIFYSPAVMNYSIP--SNVTNLEGGPGRQT 73
               :| |   ||:|:  | ||| | |  |::  : |||||:: ||    |: :::     ||:
Db          41 FNSSSPSLPVENHPICTIPSPYTDLGHDFSTLPFYSPALLGYSTSPLSDCSSV --- RQS 85

Qy          74 ISPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNREILKKKVSGNKC 133
               || : ||   |:| | : :|             | |    :|    | : | ::|:
Db          86 LSPILFWPPBSHVSSLILQQQQSRLQQNHATSGIWTEHTPHDHVEEENSKPLVKRVADIEE 145

Qy         134 ASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTI 193
               |      | | |:|||||||||||||||||||||||||||||||||||||||||||||||
Db         146 IS--VSLPGCKADMEYCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTI 203

Qy         194 DKNRKKSCQACRLRKCYEVGMVKCGSRKERCGYRLVKRQRGADEQ--LHCAGKAKRSGGH 251
               |||||||||||||||||||||:||| |||:|  |    :|| | ::  :  :|: : :|
Db         204 DKNPRKSCQACRLRKCYEVGMMKCGLRRDRCSY----QQPCAQQFPLVREFSGPMPMTGPR 259

Qy         252 APRVR----  - -ELLLDALSPEQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLTKL 301
               :  ::          |::  :||||:|:  ::|||| : |:     ||||||:|||| |
Db         260 SQEIKSIPRPLSGNEVVRISLSPEELISRIMEAEPPEIYLMKDMKKPFTEANVMMSLTNL 319

Qy         302 ADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLV 361
               |||||||||||||||||||||||||| ||| ||:|||:||||::|||||||:|||
Db         320 ADKELVHMISWAKKIPGFVELSLFDQVHLLECCWLEVLMLGLMWRSVNHPGKLIFSPDLS 379

Qy         362 LDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSM-YPLVTATQDA 420
               | |||  ||:|::|||||||| ||||||||| :|||:|||||||:|         :|
Db         380 LSRDESSCVQGLVEIFDMLLAATSRFRELKLQREEYVCLKAMILLNSNMCLGSSEGGEDL 439

Qy         421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK 480
               |  ||  ||:|||||||| |:|:|:| ||:| |||:|||||||:||  ||
Db         440 QSPRSKLLCTLDSVTDALVWAISKTGLSFQQPRTRLAHLLMLLSHIRHVSNKGMDHLTCMK 499

Qy         481 CPNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDTESPQS 520
               | :||:|||||||:||::   : | :|     | :|:||
Db         500 MKKMVPLYDLLLEMLDAHIMHSSRLSHSGPRAPAAHKDNKS 540


RESULT 6
Q8QHK9
ID   Q8QHK9       PRELIMINARY;      PRT;   565 AA.
AC   Q8QHK9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta.
GN   ERB.
OS   Paralichthys olivaceus (Flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Paralichthyidae; Paralichthys.
OX   NCBI_TaxID=8255;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kitano T., Sakimura N., Takamune K., Nagahama Y., Abe S.;
RT   "Role of estrogen receptor in gonadal sex differentiation in Japanese
RT   flounder (Paralichthys olivaceus).";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB070640; BAB85623.1; -.
KW   Receptor.
SQ   SEQUENCE   565 AA;  62521 MW;  74CAB8E1426B087D CRC64;

  Query Match             50.3%;  Score 1410.5;  DB 14;  Length 565;
  Best Local Similarity   54.5%;  Pred. No. 8.7e-122;
  Matches  291;  Conservative   66;  Mismatches  146;  Indels   41;  Gaps   12;

Qy           9 SLNSPS-SYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMIFYSPAVMNYSIPSNVTNLEG 67
               | :||  | : || |          ||| | :  |::  : || | : :|: ||    :
Db          33 SISSPGLSLDGSQPIC---------IPSPYIELGHDFATIPFYGPTIFSYAAPS    IPD 80

Qy          68 GPG-RQTTSPNVLWPTPGHLS-PLVVHR-QLSHLYAEPQKSPWCEARSLEHTLPVNRETL 124
               |   | : ||:: ||: ||:  |: :|| |       :| :||| |    :  |  | : :
Db          81 CPSVHQSLSPSLFWPSHGHMGPPMTLHPSQGPSQQGQPIQSPWGELTPRDGVL ANSKGV 139

Qy         125 KRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYI 184
               :|:          |   | | | |:|||  ||||||||||||||||||||||||||||||
Db         140 RRR--SQESEDGVVSSGGKSDLHYCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYI 197
```

```
Qy        185 CPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRKERGGRN--LVPVGKSADRQLHCA 242
              ||||||||||||||||||||||||||||||||| ||| |::   ||    ||      |   :
Db        198 CPATNQCTIDKNRRKSCQACRLRKCYEVGMTKGGMRFDHGSYRNIPKIRRLLPIESSGQPAS 257

Qy        243 GKAKRSGGHAPRVRPLLILDALSPEQLVITLIEAEPPHV-LISRPSAPFTEASMMMSLTKL 301
              |    :|     : ||   ||:||||:  ::|||||  : |:   | | |||::|||||  |
Db        258 GPKALTGPVVALMNELQPEAITPEQLIFFIMEAEPPVIYLMKLDMSGPLILANVMMSLLAL 317

Qy        302 ADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVL 361
              |||||||||:|||||||||| | ||| ||| ||:|||||||||||:|||||||:|||
Db        318 ADKELVHMITWAKKIPGFVELGLLDQVHLLECCWLEVLMMGLMWRSVDHPGKLIFSPDLIL 377

Qy        362 DRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSM-YPLVTATQDA 420
              | |:|| ||:|  ||||||:| ||| |||||  :|||:|:||||||:|      :::
Db        378 LSREEGSCVQGFSEIFDMLIAATSRVRELKLQRFEYVCLKAMILLNSNMCLSSKFGSFEL 437

Qy        421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGMEHLLNMK 480
              |  ||  ||:|||||||| |||:|:: :||  |||:|||||||:|| ||||:||  ||
Db        438 HSRSKLLLLLDAVTDALVWAIAKTGLTFRQQYTRIAHLLMILSHIRHVSNKGMDHLHCMK 497

Qy        481 CKNVVPVYDLLLEMLNAHVLRGCK---- -  SSITGSEGSPAPDSKSKEPVSQN 526
              ||:||:|||||||:||::   :           |    ||    |  ||:|
Db        498 MKNMVPLYDLLLEMLDAHIMHSSPIPHHASQQPPPTDQGFVPARDQSSQNGSSN 551


RESULT 7
Q90WS9
ID   Q90WS9      PRELIMINARY;      PRT;   592 AA.
AC   Q90WS9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ZfER-beta1 protein.
GN   ZFER-BETA1.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Menuet A., Pellegrini E., Anglade I., Blaise O., Laudet V., Kah O.,
RA   Pakdel F.;
RT   "Binding characteristics, transactivation properties and central
RT   expression of three estrogen receptor subtypes in zebrafish.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AJ414566; CAC93848.1; -.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   592 AA;  66046 MW;  7F65218E2C03F1B8 CRC64;

  Query Match             49.8%;  Score 1397.5;  DB 13;  Length 592;
  Best Local Similarity   51.6%;  Pred. No. 6.3e-121;
  Matches  300;  Conservative   79;  Mismatches  121;  Indels   81;  Gaps   16;

Qy         11 NSPSSYNCSQSILP--------LEHGSIYIPSSYVDSHHEYP----------AMTPYSP- 51
              :| :    | ::||        ::: :: ||| ||::  :|           |:| |||
Db         13 SSKADRGASPALLPRLYASPIGMDNQTVCIPSPYVFAQQDYSPPHGSFFNHGAITLYSPV 72

Qy         52 - AVMNYSIPSNVTNLPGCPGFGTTSPNVIWPTPGHL---SILVVHPQLSHLYALLGRGP 108
                ||: :  |    :|         || :||| |  |     ||      ||       |
Db         73 SSAVLGFHRPPVSESL------VPLSPTILWP-PHSLHGPPPLAYSETRSH-------SA 118
```

```
Qy        107 WCEAKSLEHTLPVNPETLP------PKVSGNPGASPVTQQPSKRQAEPGAVCSDYASGYH 160
              | ||::  ||| :  |        ::: |:  :|       | | |||||| |||||||
Db        119 WEEAKI--HILSQSSSVISHTKLLGQQLPSGNGLNPSASIVGKGDTHFGAVCHDYASGYH 176

Qy        161 YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSR 220
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||| |
Db        177 YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKSRRKSCQACRLRKCYEVGMMKGGVF 236

Qy        221 RERGGYRLVPRQR- ----SADFQLHCAGKAKR -          - -- SGGHAPRVRPLL 260
              |||| ||  | :|      |:  :  |::::              ||| |   | |
Db        237 RERGSYRGAPHRPNDQIRDDSSGGVVGIRGQSQQHLEFPISPSQHLEPSGGPA-EGPAI-- 293

Qy        261 DALSPEQLVITLLEAEPPHVLISRP-SAPFTEASMMMSLTKLADKELVHMISWAKKIPGF 319
                  ||||||  :|||||| : :  |   |:|||||||||| ||||||| ||||||||||
Db        294 -NYSPEQLVSCIIEAEPPQIYLREPVKKPYTEASMMMSITSLADKELVLMISWAKRIPGF 352

Qy        320 VELSLEDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM 379
              |||:| ||| ||| ||:::||:||||||:|||||||| ||| |:|:|| ||||:|||||
Db        353 VELTISDQVHLLECCWIDILMLGLMWRSVDHPGKLIFTPDLKLNRDEGNCVEGIMEIFDM 412

Qy        380 LLATTSRFRELKIQHKEYLCVKAMILLNSSMYPLVTAT-QDADSSRKLAHLLNAVIDALV 438
              |||||||||||||| :|||:|||||||:    :  | :| :|  |: :|||:|||||
Db        413 LLATTSRFRELKIQRFEYVCLKAMILLNSNNCSSLFGTFELVESRGKVLNLIDSVIDALV 472

Qy        439 WVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH 498
              |:|:::|:||||||:|||:|||||||:|| ||||:||| ||| |||| :||||||:|:
Db        473 WIISKTGLSSQQQSIPLAHLLMLLSHIRHLSNKGIEHLSNMKFKNVVPLYDLLLEMLDAN 532

Qy        499 VLRGCK---------SSITGSEGSPAEDSKSKEGSQNPQSQ 530
              :  :              ::  |   |       :|   :|:::
Db        533 ASQSSRMLEDRQQSPENLHTSRPQPDLKDSDQETPHSPRAE 573


RESULT 8
Q98SM7
ID   Q98SM7      PRELIMINARY;      PRT;   553 AA.
AC   Q98SM7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Akten B., Kishida M., Callard G.V.;
RT   "Estrogen receptor cDNAs in zebrafish.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF349414; AAK16742.1; -
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   553 AA;  62210 MW;  A58BBC08BL108L1L CRC64;

  Query Match             49.5%;  Score 1389.5;  DB 13;  Length 553;
```

```
  Best Local Similarity   54.4%;  Pred. No. 3.2e-120;
  Matches  283;  Conservative   73;  Mismatches  138;  Indels   27;  Gaps    9;

Qy          16 YNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIP  SNVTNLEGGPGRQT  73
               :| |   ||:|:  | ||| | |          | |      |    |: :::     ||:
Db          81 FNSSSPSLPVENHPICIPSPYTDLGTTSALCPETVPLCWGTSTSPLSDCSSV-   - RQS  85

Qy          74 TSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRC  133
               || : ||   |:| | : :|            | |    :|    | : | ::|:
Db          86 LSPTLFWPPHSHVSSLTLQQQSRLQQNHATSCTWTEHTPHDHVEEENSKPLVKRVADTEE  145

Qy         134 ASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTI  193
               |      | | |:|||||||||||||||||||||||||||||||||||||||||||||||
Db         146 TS--VSLRGKADMHYCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTI  203

Qy         194 DKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQ--LHCACKAKRSGGH  251
               ||||||||||||||||||||||:||| ||:|   |    :|| | ::  :  :|: : :|
Db         204 DKNRRKSCQACRLRKCYEVGMMKCGLRRDRSSY----QQRGAQQKRLVREFSGRMRMTGPR  259

Qy         252 APRVR------ ELLLDALSPEQLVLTLLEAEPPHV-LISRPSAPFTEASMMMSLTKL  301
               :  ::        |::  :||||:|:  ::|||||| : |:     |||||::||||| |
Db         260 SQEIKSIPRPLSGNEVVRISLSPEELISRIMEAEPPEIYLMKDMKKPFTEANVMMSLTNL  319

Qy         302 ADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLV  361
               |||||||||||||||||||||||||||| ||| |:||||:|||||::||||||:|||
Db         320 ADKELVHMISWAKKIPGFVELSLFDQVHLLECCWLEVLMLGLMWRSVNHPGKLIFSPDLC  379

Qy         362 LDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSM-YPLVTATQDA  420
               | |||  ||:|::|||||||| ||||||||| :|||:|:|||||||:|       :|
Db         380 LSRDESSCVQGLVEIFDMLLAATSRFRELKLQREEYVCLKAMILLNSNMCLGSSEGGEDL  439

Qy         421 DSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMK  480
               |  ||  :|::|||||||| |:|:|:| ||:| |||:||||||:|| ||||||:   ||
Db         440 QSRSKLLCVLDSVTDALVWAISKTGLSFQQRSTRLAHLLMLLSHIRHVSNKGMDHQHCMK  499

Qy         481 CKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPA-EDSKS  520
               | :||:|||||||:|::   : | :|     | :| ||
Db         500 MKKMVPLYDLLLEMLDAHIMHSSRLSHSGPRAPAAHKDDKS  540


RESULT 9
Q8UW75
ID   Q8UW75       PRELIMINARY;      PRT;   562 AA.
AC   Q8UW75;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta.
OS   Oryzias latipes (Medaka fish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Nobukawa T., Nakai M.;
RT   "Oryzias latipes estrogen receptor beta.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB070901; BAB79705.1; -.
DR   InterPro; IPR000194; ATPase_a/Bcentre.
DR   InterPro; IPR000445; CytC_heme_bind.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   562 AA;  62299 MW;  6ABBF9AF92FA882C CRC64;

  Query Match             48.7%;  Score 1466;  DB 13;  Length 562;
  Best Local Similarity   54.3%;  Pred. No. 4.9e-118;
  Matches  280;  Conservative   69;  Mismatches  147;  Indels   40;  Gaps   12;

Qy           1 MDIKNSPSSLNSPSSYNCSQSILPLEHGS--IYIPSSYVDSHHEYPAMIFYSPAVMNYS  57
               :|   : | : ||   | | |   | | :   | ||: | |  |:: :: ||:| : :|:
Db          19 VDSSRAGSCVLSP--NLSSSSPGLSHETSQPICIPAPYADLSHDFTSLPFYNPTIFSYA  75

Qy          58 IPSNVINLEGGPG-RQTTSPNVLWPTPGHLSPLV-VHR-QLSHLYAEPQKSPWCEARSLE  114
               ||     :  |   |: | :: | : ||: | : :|| |    : :| :||| :
Db          76 SPS---MSECPSVHQSLSASLFWQSHGHVGPTIPLHRSQARAQHGQFIQSPW----D  125

Qy         115 HTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFK  174
                 |  ::   :|          |   | | : :|||| ||||||||||||||||||||
Db         126 GVLTTSKGVRRRSQESEEA---VVSSGGKSELHYCAVCHDYASGYHYGVWSCEGCKAFFK  182

Qy         175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRL--VRRQ  232
               |||||| ||||||||||||||||||||||||||||||||||||||| ||| |:|| ||    ||
Db         183 RSIQGHTDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMTKCGVRKERSSYRSAPARRA  242

Qy         233 RSADEQLHCACKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHV-LISRPSAPFTE  291
                 |    |    ||        |     |:|||:  ::|||| : |:     | ||
Db         243 GRLTSQGRMNGPKVSSGPKESSGNFQSSLHETPEQIARMMFAEPPDIYLMKDTKKPLTE  302

Qy         292 ASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHP  351
               | :|||||  ||||||||||:|||||||||||| ||| ||| ||:|||||||||||: ||
Db         303 AVVMMSLTNLADKELVHMITWAKKIPGFVELSLLDQVHLLECCWLEVLMMGLMWRSVGHP  362

Qy         352 GKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMY  411
               |||||:||| | |:|| ||:| :|||||||:| ||| ||| || :||:|:|||||||:|
Db         363 GKLIFSPDLSLSREEGSCVQGFVEIFDMLIAATSRVRELNLQREEYVCLKAMILLNSNM-  421

Qy         412 PLVTATQ---DADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHA  468
               |:::::   :  |  ||  ||:|||||||| | |||:: :||  |||:|||||||:||
Db         422 -CLSSSEGGGELHSRSKLLCLLDAVTDALVWAIGKSGLNFRQQYTRLAHLLMLLSHIRHV  480

Qy         469 SNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCK  504
               ||||:||  || ||:||:|||||||:|||::   :
Db         481 SNKGMDHLHCMKMKNMVPLYDLLLEMLDAHIMHSSR  516


RESULT 10
Q95JC0
ID   Q95JC0       PRELIMINARY;      PRT;   250 AA.
AC   Q95JC0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ERbeta protein (Fragment).
GN   ERBETA.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDIDYMIS;
RA   Kaeppler Hanno K.G.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDIDYMIS;
RA   Telgmann R.;
RT   "Epididymal epithelium immortalized by Simian virus 40 large T
RT   antigen: a model to study epididymal gene expression.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
```

```
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY
DR   EMBL; AJ314196; CAC42097.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding, Nuclear protein, Receptor, Transcription regulation,
KW   Zinc-finger.
FT   NON_TER       1      1
FT   NON_TER     250    250
SQ   SEQUENCE   250 AA;  28072 MW;  5D946F6ORA150A77 CRC64,

  Query Match             44.4%;  Score 1243;  DB 6;  Length 250;
  Best Local Similarity   90.4%;  Pred. No. 4.4e-107;
  Matches  226;  Conservative   12;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        118 PVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSI 177
              ||||||||||||||: |||||:| |||||||||||||||||||||||||||||||||||||
Db          1 PVNRETLKRKVSGSSCASPITSPSSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFRRSI 60

Qy        178 QGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYPLVPRQRSADF 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||:|:||
Db         61 QGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYPVMRFQKSSDF 120

Qy        238 QLHCAGKAKRSGGHAPFVKPLLIDALSPFQLVLTLLEAEPPHVLISRPSAIPFLEASMMMS 297
              |||| |||::||| ||:|||| ||||||||||||||||||:|||| ||||||||||
Db        121 QLHCISKAKKNGGHVTFVKPLLISALSPFQLVLTLLEAEPPHVLVSRPSTPFTEASMMMS 180

Qy        298 LTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAP 357
              |||||||||||| |||||||||||| ||||||||||||||:|||||||||||||||||||
Db        181 LTKLADKELVHMIGWAKKIPGFVELSLCDQVRLLESCWMEVLMVGIMWRPSIHPGKLVFA 240

Qy        358 PDLVLDRDEG 367
              ||||||||||
Db        241 PDLVLDRDEG 250


RESULT 11
Q8UWA9
ID   Q8UWA9      PRELIMINARY;      PRT;   581 AA.
AC   Q8UWA9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor.
GN   ER.
OS   Cnemidophorus uniparens (Whiptail lizard).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Teiioidea;
OC   Teiidae; Cnemidophorus.
OX   NCBI_TaxID=37197;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21490797; PubMed=11604222;
RA   Sumida K., Ooe N., Saito K., Kaneko H.;
RT   "Molecular cloning and characterization of reptilian estrogen receptor
RT   cDNAs.";
RL   Mol. Cell. Endocrinol. 183:33-39(2001).
DR   EMBL; AB055221; BAB79437.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq
DR   InterPro; IPR001292; Oestrgn_receptor
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.

DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   581 AA;  65650 MW;  A99P6C844PCF0D58 CRC64;

  Query Match             44.0%;  Score 1235.5;  DB 13;  Length 581;
  Best Local Similarity   49.1%;  Pred. No. 6.7e-106;
  Matches  265;  Conservative   78;  Mismatches  122;  Indels   75;  Gaps   14;

Qy         10 LNSFSSYNCSQSTLFLEHGSTYTESSYVTDSH    HEYT          SAMTTYSTAVMNY 56
              || |      | :|||     |   ||||:      ||          |   || | ::|
Db         26 LNRP------QLKIPLERP---ISEMYVDSNKTGVFNYPEGATYDFSTAAPVYSSASLSY 76

Qy         57 SIPGNVTNLFQGGPG---------PQTTSPNVLWPTPGHLSPIVVHSQLS---HLYAEPQK 104
              :    ||   | |                || |   |   ||| : |       :|  ||
Db         77 A----STNESFGSGNLGGLHSLNNVPPSPVVFLQTAPQLSPFTHHHINQQVPYYLENEPSS 132

Qy        105 SPWCFAASLEHTLFVNFETLKPKVSGNRCASPVTGP---GSKRDAHFCAVCSDYASGYHY 161
              |  ||      | :   : :  | | ::   |     | ::  :||||||||||||
Db        133 SAMREAFPIAFYKPGS----ENRHHGGKASNSEKGSLSMESIKLRYCAVCNDYASGYHY 188

Qy        162 GVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRR 221
              ||||||||||||||||||:|||||||||||||||||||||||||||||:| | |:
Db        189 GVWSCEGCKAFFKRSIQGHNDYMCPAINQCIIDKNRRKSCQACRLRKCYEVGMMKGGIRK 248

Qy        222 FRGGYPIVPRQRSADFQLHCAGKAKRSGGHAPFVPF----------LIID---------A 262
              :| | |::: :|  ||           | :|  | |          |::          :
Db        249 LRKGGPMIFKPKQPDF- - -----LDGRNAVAVTPARNTTIWPSPLMIKHSKKNSPALS 299

Qy        263 LSPFQLVLILLEAEPPHVLISRPSAPPILEASMMMSLIKLADKELVHMISWAKRIPGFVE 321
              |: |||:|  |||:|||| |     ||:|||:|||:|  || |||:|||||:|||:||||
Db        300 LTAFALKVSALIEAELFIVYSEYGFSSFFSEASVMTLLTNLADRELVHMITWAKRVPGFVD 359

Qy        322 LSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLL 381
              |:| ||| |||  |:|:||:||:|||:|:|||:|||:|:||| :| |||| :|||||||
Db        360 LALHDQVHLLECAWLEILMIGLIWRSLEHPGKLIFAPNLILDRSQGMCVEGFVEIFDMLL 419

Qy        382 ATTSRFRFLKIQNKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWV 440
              ||:||| : :| :|:|:|:|:|||| :|  :::| :  :   :  :|: :|| | :
Db        420 ATSSRFRMMNIQGEEFVCLKSIILLNSGIYTFLSSTLRSLEEKEHIHRVLDKITDILIHL 479

Qy        441 IAKSGLSSQQQSMRLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
              :||||:| |||  ||| ||::|||:|| |||||||| ||||||||:||||||:|| |
Db        480 MAKSGLSLQQQHRRLAQLLLMLSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEMLDAHRL 539


RESULT 12
Q8UWB0
ID   Q8UWB0      PRELIMINARY;      PRT;   587 AA.
AC   Q8UWB0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor.
GN   ER.
OS   Caiman crocodilus (Spectacled caiman) (Caiman sclerops).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Crocodylidae; Alligatorinae; Caiman
OX   NCBI_TaxID=8499;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21490797; PubMed=11604222;
RA   Sumida K., Ooe N., Saito K., Kaneko H.;
RT   "Molecular cloning and characterization of reptilian estrogen receptor
RT   cDNAs.";
RL   Mol. Cell. Endocrinol. 183:33-39(2001).
DR   EMBL; AB055220; BAB79436.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   Pfam; PF00105; zf-C4; 1.
```

```
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   587 AA;  66514 MW;  DE17F28847FCF513 CRC64;

  Query Match             43.4%;  Score 1216.5;  DB 13;  Length 587;
  Best Local Similarity   49.9%;  Pred. No. 4.9e-104;
  Matches  260;  Conservative   81;  Mismatches  127;  Indels   53;  Gaps   13;

Qy          23 LPLEHGSIYIPSSYVDSH----HEYPAMTFY--SPAVMNYSIPSNVTNLEGGPGRQT--- 73
               :||:     :   ||:::      ||  | |  : |   ||     |:|   |  ::
Db          33 IPLDRS--LSEMYVENNKTGIFNYPEGTTYDFATAAPVYS----STSLSYAPTSESYGS 85

Qy          74 -----------TSPNVLWPTPGHLSPLVVHR-QLSHLYAEPQKSPW-CEAPSLEHTL 117
                          || |   |  ||| | | |    | |  :| :   || |
Db          86 SSLGGFHSLNNVPPSPVVFLQTAPQLSPFVHHHSQQVPYYLENDQSGFGMREAASSTFYR 145

Qy         118 PVNRETLKRKVSGNRCASPVTGPG-----SKRDAHFCAVCSDYASGYHYGVWSCEGCKAF 172
               |     |  ||    |  :      | ::  :||||:||||||||||||||||||||||
Db         146 P--  SADSRHQSGRERMSSTSEKASLSMESTKETRYCAVCNDYASGYHYGVWSCEGCKAF 202

Qy         173 FKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQ 232
               ||||||||||:|||||||||||||||||||||||||||||||:| | |::| | |:::::
Db         203 FKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRMLKQK 262

Qy         233 RSADEQLHCAGKAKRSGGHAPRV  RELLLD--------ALSPEQLVLTLLEAEPPHVL 281
               |  :||    |:   :    | :    |::           :|: ||:|  |||||||  |
Db         263 RQREEQDARNGETATAEMRTPTLLWTSPLVTKHTEKNSPAISLIAFQMVSALTKAEPPQVY 322

Qy         282 IS-RPSAPETEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLM 340
               |: || |||||  || |||:|||||:|||::||||:|:| ||| |||  |:|:||
Db         323 SEYDPNRPFNEASMMTLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILM 382

Qy         341 MGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCV 400
               :|:::||::|||||:|||:|:|||::|||||::|||||||| :||| : || :|:::|:
Db         383 IGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATAARFRMMNLQGEEFVCL 442

Qy         401 KAMILLNSSMYPLVTATQDADSSRKLAH-LLNAVTDALVWVIAKSGI 446
               |::|||||  :|  :::|   :   :  | :|: :|| |: :|||||:| |||  ||| ||
Db         443 KSIILLNSGVYTFLSSTLKSLEEKDYIHRVLDKITDTLIHLMAKSGLSLQQQHRRLAQLL 502

Qy         460 MLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVL 500
               ::|||:|| ||||||||  ||||||||||:||||||-|:|| |
Db         503 LILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEMLDAHRL 543


RESULT 13
Q98SM9
ID   Q98SM9       PRELIMINARY;      PRT;   569 AA.
AC   Q98SM9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor alpha.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Akten B., Kishida M., Callard G.V.;
RT   "Estrogen receptor cDNAs in zebrafish.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF349412; AAK16740.1;  -.
DR   HSSP; P03372; 1HCP.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   569 AA;  62846 MW;  CFA0C1AD4476A54C CRC64;

  Query Match             43.1%;  Score 1210;  DB 13;  Length 569;
  Best Local Similarity   49.0%;  Pred. No. 1.5e-103;
  Matches  277;  Conservative   71;  Mismatches  129;  Indels   88;  Gaps   19;

Qy          26 EHGSIYIPSS--YVDSHHEY--PAMTF--YSPA----VMNYSIP-----SNVTNLEGGPG 70
               || :  | ||  |:|  :||  |  ||   |||     |  |  |       ::  | ||
Db           6 EHSAGGISSSVNYILDGAYEYPNPTQTFGTSSPAEPASVGYYPAPPDPHEEHLQTLGGG-- 63

Qy          71 RQTTSPNVLWPTPGHLSPLVVH-----  --RQLSHLYAEPQKSP---------------W 107
                 ::|| :  |:   ||| : |          |:|: | :   |
Db          64 --SSSPLMFAPSSPQLSPYLSHHGCHHTIPHQVSY-YLDSSSSIVYRSSVVSSQQAAVCL 120

Qy         108 CEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCE 167
               ||    |     :|: |    :|:| |     |  | ::  ||||||||||||||||||
Db         121 CE----ELCSATDRQEL - YTCSRAAG---GFDSGRETPFCAVCSDYASGYHYGVWSCE 170

Qy         168 GCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYR 227
               |||||||||||||||:||||||||||:|||||||||||||||||||:| | |::| | |
Db         171 GCKAFFKRSIQGHNDYVCPATNQCTIDPNRRKSCQACRLRKCYEVGMMKGGIRKDR GGR 229

Qy         228 LVRRQR           SADEQLHCAG          KAKRSGGHAPRVRELLLDALSPEQL 268
                |||:|           |: :|   ||          : |||||    :        :|||:|
Db         230 SVRREPPPSSNEDPDKSSSDQCSRAGVRTTGPQDERRERRSGGVVSIL-------GMSPDQV 283

Qy         269 VLTLLEAEPPHVLI SETSAPTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQ 327
               :| || |||| |    : | :|:|| :||  || :|||||||||:||||:||| |||| ||
Db         284 LLLLLGAEPPAVCSRQKHSRPYTEITMMSLLTNMADKELVHMIAWAKKVPGFQDLSLHDQ 343

Qy         328 VRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRF 387
               |:||||  |:|||:||:||||  |||||||  ||:||| ||:|||: |||||||||  :||
Db         344 VQLLESSWLEVLMIGLIWRSIHSPGKLIFAQDLILDRSEGECVEGMAEIFDMLLATVARF 403

Qy         388 RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAH-LLNAVTDALVWVIAKSGI 446
               | |||: :|:::|:||:|||:||  :   ::  :      :  :|: :||||:: |:|||
Db         404 RSLKLKLEEFVCLKAIILLNSGAFSFCSSPVEPLMDNFMVQCMLDNITDALIYCISKSGA 463

Qy         447 SSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAH- ---- V 499
               | | || | | ||:||||:|| ||||||||  |||| ||:|||||||||||:|      |
Db         464 SLQIQSEPQAQLLLLLSHIRHMSNKGMEHLYRMKCKNFVPLYDLLLEMLDAQRFQSSGKV 523

Qy         500 LRGCKSSITGSECSPAEDSKSKEGS 524
                |    |       :|   | |   |
Db         524 QRVWSQSEKNPPSTPTTSSSSSNNS 548


RESULT 14
Q90WV1
ID   Q90WV1       PRELIMINARY;      PRT;   564 AA.
AC   Q90WV1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor alpha.
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
```

```
OC   Cyprinidae; Carassius.
OX   NCBI_TaxID=7957;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Choi C.-Y., Habibi H.R.;
RT   "Molecular Cloning and Expression Pattern of the Estrogen Receptor
RT   alpha in the Goldfish (Carassius auratus).";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY
DR   EMBL; AY055725; AAL12298.1; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   564 AA;  62763 MW;  EC34E1C0B6631382 CRC64;

  Query Match             42.4%;  Score 1188;  DB 13;  Length 564;
  Best Local Similarity   49.6%;  Pred. No. 1.6e-101;
  Matches  263;  Conservative   72;  Mismatches  119;  Indels   76;  Gaps   15;

Qy         26 EHGSIYIPSS--YVDSHHEY-------------PAMTFYSPAVMNYSIP---SNVTNLE 66
              ||  :  | ||  ||:|  :::             ||   | ||  :   |     ::  ||
Db          6 EHSTGGISSSLNYILDGAYDFSNLSEIFGTSSPIEPASFGYYPAPPDPHAPPAEEHLQNLG 65

Qy         67 GGPGRQTTSPNVLWPTPGHLSPLVVH-------RQLSHLYAEPQKSPW------------ 107
               |||    || :  |:   ||| : |       |:|:    |  : :
Db         66 TGPG----SPLMFAPSSPQLSPYLNHHGGHHSTHQVSYYLDSPSSTVYRSSVVCSQQAGV 121

Qy        108 --CEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWS 165
                ||    |     :|: |    ||:| |     |  |::::  ||||||||||||||||||
Db        122 GICP    EICSAADRQEL---FSGSRAAG---GFDSEKETRFCAVCSDYASGYHYGVWS 171

Qy        166 CEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRERGG 225
              ||||||||||||||||||:|||||||||:|||||||||||||||||:| | |:|| |
Db        172 CPGCKAFFKRSIQGHNDYVCPATNQCTIDRNRRKSCQACRLRKCYEVGMMKGGIRKERRG 230

Qy        226 YPLVPRQP----------SADEQLHCAG------KAKRSGCHAPRVRELLLDALSPEQLV 269
              | :|| |          |  || :  |      : :|       |  |    : |:|::
Db        231 GPAIRPDPRPSGNEDRVKSYSEQSNRTGIPTAFPQDPRKSSSTGVVSIL---CMPPQQVL 287

Qy        270 LTLLEAEPPHVLI-SRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQV 328
              . || |||| |    : | |:|| :|  ||  :|||||||||:|||:|||  :||| |:|
Db        288 VLIIGAFPPAVCSRQKHSPPYTEITMMSLITNMADKELVHMIAWAKRVPGFQDLSLHDQV 347

Qy        329 RLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFR 388
              :|||| |:|||||:||:|||  |||||||| ||:|||:||:|||| |||||||||| :|||
Db        348 QLIESTWLEVLMIGLIWRSIHSPGKLIFAQDLILDRNEGECVEGMAEIFDMLLATVIRFR 407

Qy        389 ELKLQHKEYLCVKAMILLNSSMYPLVTA-TQDADSSRKLAHLLNAVTDALVWVIAKSGIS 447
              |||: :|:::|:||:|||||  :   ::  :    |  :  :|: :||| :: |:||| |
Db        408 SLKLKLEEFVCLKALILLNSGAFSFCSSPVEPLMDSFMVQCMLDNTIDAEIYGISKSGAS 467

Qy        448 SQQQSMRLANLIMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
              | || | | ||:|||||:|| |||||||  |||| ||:|||||||:|
Db        468 LQLQSRRQAQLLLLLSHIRHMSNKGMEHLYRMKCKNRVPLYDLLLEMLDA 517


RESULT 15
Q8QHL0
ID   Q8QHL0      PRELIMINARY;      PRT;   578 AA.
AC   Q8QHL0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor alpha.
GN   ERA.
OS   Paralichthys olivaceus (Flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Paralichthyidae; Paralichthys.
OX   NCBI_TaxID=8255;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kitano T., Sakimura N., Takamune K., Nagahama Y., Abe S.;
RT   "Role of estrogen receptor in gonadal sex differentiation in Japanese
RT   flounder (Paralichthys olivaceus).";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB070629; BAB85622.1; -.
KW   Receptor.
SQ   SEQUENCE   578 AA;  63651 MW;  3AA5931296A6A714 CRC64;

  Query Match             41.6%;  Score 1166.5;  DB 13;  Length 578;
  Best Local Similarity   48.4%;  Pred. No. 1.7e-99;
  Matches  259;  Conservative   79;  Mismatches  136;  Indels   61;  Gaps   16;

Qy         36 YVDSHHEY----PAMT---------PYSPAVMNYSIPSNVT--NLEGGPGRQTTSPNVLW 80
              :::  ::|    || |         :|| :  :  ||: :  :|  ||    |||:|
Db         17 FLEGIYDYAAPTPAQTPLYSHSTSGYYSAPLDAHGPPSDGSRHSIGSGP----TSPHVYV 72

Qy         81 PTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVNRE------TLKRKVSGNRCA 134
              |:   |||  :     || | |   :     || :   || ||           |     |
Db         73 PSSPRLSPFM--HPPSHHYLETTAT--SVYRSSQQ--PVTREDHCGPRDESFSVGETGAA 126

Qy        135 SPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTID 194
              :   |    ::  ||||||||||||||||||||||||||||||||||||:|||||||||
Db        127 AGAEGFFMAKETRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTID 186

Qy        195 KNRRKSCQACRLRKCYEVGMVKGGSRRERGGYRLVR---------RQPSADEQLHCA--- 242
              :|||||||||||||||||||:| | |::|  :| | |                | ::: :| |
Db        187 PNRRKSCQACRLRKCYEVGMMKGGVRKDRRGGRNDRDRKASKIVDHKTVPL 245

Qy        243 -----GKAKRSGCHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP--SAPFTEASMM 295
                    :  :|| :       : |: |:|::: |  |||| :| ||   : :||| :||
Db        246 QDGRKSSSSTAGGKSS------VTAMLPDQVLVLLQGAEPP-ILCSRQKLNQPYTEVIMM 298

Qy        296 MSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLI 355
               || :||:|||||:||||:||::||| |||:|||| |:||||:||:||||  ||||
Db        299 TLLTSMADRELVHMIAWAKKLPGFLQLSLHDQVQLLESSWLEVLMIGLIWRSIHCPGKLI 458

Qy        356 FAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVT 415
              || ||:|||:|| |||: |||||||| |||| ||| :|: |:|:|:|||  :   |
Db        359 FAQDLILDRNEGNCVEGMAEIFDMLLATASRFRMLKLKSFEFFCLKAIILLNSGSFSFCT 418

Qy        416 ATQD-ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMILSHVRHASNKGME 474
              | :   ::  :  :|  :||||:  |::||   |||  | | ||:|||||:|| ||||||
Db        419 GTMEPLHNTAAVQDMLFTITDALIHHISQSGCPVQQQWERQAQLIILLSHIRHMSNKGME 478

Qy        475 HLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSPEGSQNPQS 529
              || :||||| ||:|||||||||:|| |   : :         |:   |  |: |  |
Db        479 HLYSMKCKNKVPLYDLLLEMLDAHCLH--RPARPAQSWLQADREPSAAGNNNNNS 531


RESULT 16
Q95L13
ID   Q95L13      PRELIMINARY;      PRT;   431 AA.
AC   Q95L13;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor alpha (Fragment).
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ENDOMETRIUM;
```

```
RA   Ing N.B.;
RT   "Estrogen receptor alpha mRNA is stabilized by estradiol in an
RT   endometrial adenocarcinoma cell line.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC    !  SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC    !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AY033393; AAK52104.1; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestran_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1       1
SQ   SEQUENCE   431 AA;  48694 MW;  BAFB6CB88D7B1E99 CRC64;

  Query Match             41.6%;  Score 1165.5;  DB 6;  Length 431;
  Best Local Similarity   55.4%;  Pred. No. 1.4e-99;
  Matches  229;  Conservative   69;  Mismatches   66;  Indels   49;  Gaps    8;

Qy        142 SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSC 201
              |  ::   :|||||:|||||||||||||||||||||||||||||||:|||||||||||||
Db         14 SAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSC 73

Qy        202 QACRLRKCYEVGMVKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGC             250
              |||||||||||||:| | |::| | |::: :|  |:       :|  ||
Db         74 QACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDG-EGRNEAVPSGDMRATNLWPSPL 132

Qy        251    -HAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSA--PFTEASMMMSLTKLADKE 305
                 |  :   :|  :|: :|::  |||||| :: :|      ||:|||||  || |||:|
Db        133 MIKHTKKNSPVL--SLTADQMISALLEAEPP-IIYSEYDLPDPFSEASMMGLLTGLADRE 189

Qy        306 LVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRD 365
              |||||:||||:||||:|:| ||| |||  |:|:||:||.|||.:|||||:|||:|:|||:
Db        190 LVHMINWAKKVPGFVDLALHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRN 249

Qy        366 EGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRK 425
              :|||||::|||||||||:|||| : || :|:::|:|::|||||:|  :::|  :   :
Db        250 QGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKD 309

Qy        426 LAH LLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNV 484
                | :|: :|| |: ::|||:|: |||  |||  |:|||| || |||||||| :||||||
Db        310 HIHRVLDKITDFLIHLMAKAGLTLQQQHRRLAQFLLILSHIRHMSNKGMEHLYSMKCKNV 369

Qy        485 VPVYDLLLEMLNAHVLR  - - - - - - - - - - - -GCKSS    ---ITG 509
              ||:|||||||||:|| |                        || ||       |||
Db        370 VPLYDLLLEMLDAHRLHAPANPGSTPPEDVNQSQLATTGCTSSHSLQTYYITG 422


RESULT 17
Q90WH6
ID   Q90WH6       PRELIMINARY;      PRT;   620 AA.
AC   Q90WH6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor alpha.
OS   Clarias gariepinus (Sharptooth catfish) (African catfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC   Clariidae; Clarias.
OX   NCBI_TaxID=13013;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PITUITARY;
RA   Jacobs J.J.L., Boqerd J.;
RT   "Random 3' RACE: a modification of 3' RACE in order to partially
RT   clone (large) cDNAs.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.

CC    !  SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC    !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; X84743; CAA57560.1; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001292; Oestran_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02159; Oest_recep; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   620 AA;  68043 MW;  8E7767DB1982A659 CRC64;

  Query Match             41.2%;  Score 1152;  DB 13;  Length 620;
  Best Local Similarity   49.1%;  Pred. No. 1.4e-98;
  Matches  255;  Conservative   76;  Mismatches  138;  Indels   50;  Gaps   13;

Qy         28 GSIYIPSSYVDSHHEY- --- PAMTFYSPAV-MNYSIPSNVTNLEGGPGRQTTSPNVLW 80
              | :   : |:|   :|      |:: :|| |  :  :       ||:      : || |
Db         47 GGMSSTAHYLDGTFDYASNPPNPSVDYYSAAPDLQVAPEPQEENLQPLGANGSASPMVFV 106

Qy         81 PTPGHLSPLVVH-------RQLSHLYAEPQKSPWCEARSL------ EHTLPVNRETLKR 126
              :  ||| : |          |   | || :|   :  |       |      :| : :
Db        107 SSSPQLSPFLSHPPAGQHVAQQVPYYLEPSGTPVYRSSVLASAGSRVELCSATSRQDVYT 166

Qy        127 KVSGNR--CASPVTGP-GSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDY 183
              |  .|   ||  .|  |  .:  :|||||||||||||||||||||||||||||||||||
Db        167 AVGASRPSGASGISGATGLVKEIRYCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDY 226

Qy        184 ICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYR---------LVRRQRS 234
              :|||||||||:|||||||||||||||||||:| | :|| |           |  |:|
Db        227 VCPATNQCTIDPNRRKSCQACPLRKCYEVGMMKGGYPEERGGPAMKHSRPHGGLKERERG 286

Qy        235 ADEQLHCAGKAKR---------SGGHAPRVREFLLDALSPFQLVETLLEAEPPHVLISRP 285
              :|    :|   |         : |    |  ::   ::|||:| || |||  | ||
Db        287 YNEAQ--SGSDARFAPPQDGRSISGTGSAVAGVV--CMAPEQVLLLLLRAEPP-TLSSRQ 341

Qy        286   SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGL 343
                | |::| ::|  || :||:||||||:||||:||| :||| |||:|||| |:|:||:||
Db        342 KLSSPFYSELIMMSLILNMADRELVHMIAWAKKVPGFQDLSLHDQVQLLESSWLEILMIGL 401

Qy        344 MWRSIDHPGKLIFAPDLVLDRDEGECVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAM 403
              :|||| |||||||| ||:||| ||:|||: |||||||| :||| ||| :|:::|:||:
Db        402 IWRSIHSPGKLIFAQDLILDRTEGECVEGMAEIFDMLLATVARFRALKLKSEEFVCLKAI 461

Qy        404 ILLNSSMYPLVIA-TQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLL 462
              |||||| :   ::  :       :  ::: :|||||  |::|||  | || | | |:|||
Db        462 ILLNSSAFSFYSSPVEPLRDGFMVQCMMDNIIDALVHYISQNGIPVQLQSRRQAQLLLLL 521

Qy        463 SHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLR 501
              ||:|| | |||||| |||||| ||:|||||||||:|| ||
Db        522 SHIRHMSYKGMEHLYNMKCKNKVPLYDLLLEMLDAHRLR 560


RESULT 18
Q90ZM8
ID   Q90ZM8       PRELIMINARY;      PRT;   554 AA.
AC   Q90ZM8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor (Fragment).
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21244662; PubMed=11831759;
RA   Thornton J.W.;
RT   "Evolution of vertebrate steroid receptors from an ancestral estrogen
```

```
RT   receptor by ligand exploitation and serial genome expansions.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:5671-5676(2001).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AY028456; AAK20929.1; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1      1
FT   NON_TER     554    554
SQ   SEQUENCE   554 AA;  61324 MW;  9E7ACBB7933DF47F CRC64;

  Query Match             39.9%;  Score 1118;  DB 13;  Length 554;
  Best Local Similarity   49.5%;  Pred. No. 5.1e-95;
  Matches  248;  Conservative   64;  Mismatches  135;  Indels   54;  Gaps   13;

Qy         35 SYVDSHHEYPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQ 94
              ||:    | |     | |      | ||   ::| | ||    |:   | |
Db         76 SYLPPSTELPQ---YVP-----SSPSAPYSMELGAGR----PHGYDPGP---------- 112

Qy         95 LSHLY---AEPQKSPWCEAASL---EHTLPVNRPTIKRKVSGNRCASPVTGPGSKRDAHF 148
                 ||    |    |: | : :       :     |  | ||     |  | |  :|   |
Db        113 -QSLYPGGVPSSAPPYSFQGQVVGGGGAMSAMCITEPRHVSSGSLPSS-TPP--ERSTQF 169

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              |||||||||||||||||||||||||||||| |||||:|||||||||||:|||||||||||
Db        169 CAVCSDYASGYHYGVWSCEGCKAFFKRSTQGHNDYMCPATNQCTIDRNRRKSCQACRLRK 228

Qy        209 CYEVGMVKGGSRRERGGYRIVRPQPSADEQLHCAGKAKRSGCH------------APPV 255
              |||||||| | |::| |:| |: :|         |:    ||                ||
Db        229 CYEVGMVK-GVRKDRKGFPGVKHKPKRPIPQENGGEGGAGGQQPVSETFPQGEEPSGPED 287

Qy        256 RELLLDALSPEQLVLTLLEAEPPHVLIS-RPSAPFTEASMMMSLTKLADKELVHMISWAK 314
              ||  . .|  .|.. |||:||| || |  |  | |||| |  :|| ||| :|||||:|||
Db        288 RESAVSSLEADQVTSALLEAEPPTVLSSYDPDKPVTEASLMAALTSLADRELVHMITWAK 347

Qy        315 KIPGFVELSLFDQVRLLESCWMFVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGIL 374
              |||||  : | |||:||| ||:|:|:::||||||| ||:| |||:|:| |:: : |||:|
Db        348 KIPGFTAIGLSDQVQLLECCWLEILIVGLIWRSIDRPGQLIFAPNLILGREDARNVEGML 407

Qy        375 EIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSM-YPLVTATQDADSSRKLAHLLNAV 433
              ::|||| | |||||  |: :|||:|||||||| : ' |  '  '  '  '  '  '  '|  |
Db        408 DMFDMLLVTVSKFRELHLRREEYVCLKAMILLNSGVFPCLSNSAGEQTNVQLIQQILEKV 467

Qy        434 TDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLE 493
               |||   |     |  | | ||: ||:||| :|| ||||:||| :|| |||:|:|||||
Db        468 MDALQSTIGHIEASPPQHSRRLSQLLLLLSQIRHISNKGIEHLNSMKRKNVIPLYDLLLE 527

Qy        494 MLNAHVLRGCKSSITGSECSP 514
              :|:|| |:      ||   ||
Db        528 LLDAHSLQN-----TGLRISP 543


RESULT 19
Q9DDJ3
ID   Q9DDJ3      PRELIMINARY;      PRT;   458 AA.
AC   Q9DDJ3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor (Fragment).
GN   ER.
OS   Halichoeres trimaculatus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidei;
OC   Labridae; Halichoeres.
OX   NCBI_TaxID=147232;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim S.J., Kei O., Takemura A., Nakamura M.;
RT   "Partial sequence and expression of androgen and estrogen receptor
RT   genes in the protogynous wrasse, Halichoeres trimaculatus.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF326201; AAG48341.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   458 AA;  51319 MW;  019482382111F277 CRC64;

  Query Match             39.2%;  Score 1099;  DB 13;  Length 458;
  Best Local Similarity   56.2%;  Pred. No. 2.3e-93;
  Matches  230;  Conservative   64;  Mismatches   85;  Indels   30;  Gaps   10;

Qy        144 RDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQA 203
              ::  :|||||||||||||||||||||||||||||||||||||||:|||||||||:|||| ||
Db         23 KFMRYCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDRNRRKSYQA 82

Qy        204 CRLRKCYEVGMVKGGSRRERGGYRLVRPQRSADEQLHCA-----GKAKR--------SGG 250
              |||||||||:|.| |:|..|   |::||     |::  |      |  :|        :||
Db         83 CRLRKCYEVGMMKGGARKDR--GRVLRR----DKRRTGTSDKDKGSKERDERTAPPQAGG 136

Qy        251 HAPPVRELLLDALS--PEQLVLTLLEAEPPHVLISRP--SAPFTEASMMMSLTKLADKEL 306
              :  | |     :|  |:|::| |  || | :| ||   | |:|| :||  || :||:||
Db        137 NTATVWEENPQWISMPPDQVLLLLQGAETP-ILYSRQKLSRPYTEVTMMTLLISMADREL 195

Qy        307 VHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDE 366
              ||||:|||:|||::|:  |||:|||| |:||||:||:|||  || ||| |:|||:|||:|
Db        196 VHMIAWAKKLPGFLQLTHHDQVQLLESSWLEVLMIGLIWRSIHCRGKHIFAQDLILDRNE 255

Qy        367 GKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQD-ADSSRK 425
              | |||: |||||||||| || ||| :|::|:||::||||  :  | | :   |
Db        256 GDCVEGMAEIFDMLLATTSPFRMLKLKPEEFVCLKAIVLLNSGAFSFQTGTMEPLHDSAP 315

Qy        426 LAHLLNAVIDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVV 485
              :  :|: :|||||  |::|| |: ||| | | ||:||||:|| |||||||| :|||| |
Db        316 VQDMLDLITDALIHHISQSGCSAHQQSRFQAQLLLLLSHIRHMSNKGMEHLYSMKCKNKV 375

Qy        486 PVYDLLLEMLNAHVL----RGCKS-SITGSECSPAEDSKSKEGSQNPQS 529
              |:|||||||:|| |     |  :| | |  | :  :   :   | |  |
Db        376 PLYDLLLEMLDAHRLHRPDRPAESWSQTDGEPAYSSSATTINDSNNNSS 424


RESULT 20
Q98SM8
ID   Q98SM8      PRELIMINARY;      PRT;   491 AA.
AC   Q98SM8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor gamma (Fragment).
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Akten B., Kishida M., Callard G.V.;
RT   "Estrogen receptor cDNAs in zebrafish.";
RL   Submitted (FEB 2001) to the EMBL/GenBank/DDBJ databases.
CC   !  SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   !  SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF349413; AAK16741.1;  -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1       1
SQ   SEQUENCE   391 AA;  44601 MW;  45E981266A815E91 CRC64;

  Query Match             37.8%;  Score 1061.5;  DB 13;  Length 391;
  Best Local Similarity   58.2%;  Pred. No. 5.6e-90;
  Matches  219;  Conservative   55;  Mismatches   67;  Indels   35;  Gaps    7;

Qy        186 PATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQR----- SADEQL 239
              |||||||||||:|||||||||||||||||||:||| ||||| || | :|      |:   :
Db          1 PATNQCTIDKSRRKSCQACRLRKCYEVGMMKCGVRRERCSYRGAPRPRNPQIRDSSGGVV 60

Qy        240 HCAGKAKR ------- - -SGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRP 285
              |::::              ||| |   | |     |||||  :|||||| : :  |
Db         61 GLRGQSQQHLEFPLSPSQHLFPSGGRA EGRGL --NYSPEQLVSCLLEAEPPQTYLREP 116

Qy        286 -SAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLM 344
               :||||||||||| ||||| | ||||||||||||||:| ||| ||| ||:::||:|||
Db        117 VKKPYTEASMMMSLTSLADKEPVLMISWAKKIPGFVELTLSDQVHLLECCWLDILMIGLM 176

Qy        345 WRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMI 404
              |||:|||||||| ||| |:|:|| |||||:|||||||||||||||||| :|||:|:|||
Db        177 WRSVDHPGKLIFTPDLKLNREEGNCVEGIMEIFDMLLATTSRFRELKLQREEYVCLKAMI 236

Qy        405 LLNSSMYPLVTAT QDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLS 463
              ||||: :  :  | :| :|  |: :|||::|||||||:|:::|:|||||:|||:|||||
Db        237 LLNSNNWSSLPQTPEDVESRGKVLNLLDSVTDALVWIISRTGLSSQQQSIRLAHLLMLLS 296

Qy        464 HVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCK --------SSTIGSECSP 514
              |:|| ||||:||| ||| |||| :|||||||||:|:  :  :           ::  |   |
Db        297 HIRHLSNKGIEHLSNMKKKNVVLLYDLLLEMLDATLQSSRMLEDRQQSPENLHTSRPQP 356

Qy        515 AEDSKSKEGSQNPQSQ 530
                :|   :|:::
Db        357 DLKDSDQETPHSPRAE 372


RESULT 21
Q9DDZ4
ID   Q9DDZ4       PRELIMINARY;       PRT;   335 AA.
AC   Q9DDZ4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor subtype 1 (Fragment).
OS   Micropterus salmoides (Largemouth bass).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
OC   Centrarchidae; Micropterus.
OX   NCBI_TaxID=27706;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bowman C.J., Sabo-Attwood T.L., Denslow N.D.;
```

```
RT   "Largemouth Bass Estrogen Receptor Subtype 1 Partial cDNA Sequence.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF253062; AAG44622.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1       1
FT   NON_TER     335     335
SQ   SEQUENCE   335 AA;  37768 MW;  40EBAE950000497F CRC64;

  Query Match             33.6%;  Score 942.5;  DB 13;  Length 335;
  Best Local Similarity   56.4%;  Pred. No. 5e-79;
  Matches  193;  Conservative   51;  Mismatches   59;  Indels   39;  Gaps    7;

Qy        157 SGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVK 216
              || |||||||||| ||||||||||||:|||||||||:|||||||||||||||||||:|
Db          2 SGSHYGVWSCEGCKPFFKRSIQGHNDYMCPATNQCTIDRNRRKSCQACRLRKCYEVGMMK 61

Qy        217 CGSRRERCGYRLVRRQRSA---- ---------- - -DEQLHCAGKAKRSGGHAPRVR 256
              | |::| |  | | :| |                    | : | :  |   || :
Db         62 GGVRKDK-GRVLRRDKRKAGTNDRDKASKDLEYRIVPPQDKRKHSSSSAGGGGKSS--- 117

Qy        257 FLTIDAISPFQLVTTLFAFPPHVLISRP--SAPFTFASMMMSLTK ---- --LADKEL 306
                :  :||:|::| |  |||| :| ||   | |:|| ::|  ||         :||||
Db        118 ---VTGMSPDQVLLLLQGAEPP-MLCSRQKLSRPYTEVTIMTLLTSMADKELVHMADKEL 173

Qy        307 VHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDE 366
              |||:|||:|||::||| |||:|||| |:||||:||:|||  |||||| ||:|||:|
Db        174 VHMITWAKKLPGFIQLSLHDQVQLLESSWLEVLMIGLIWRSIDHPGKLIFAQDLILDRNE 233

Qy        367 GKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQD-ADSSRK 425
              | |||| :||||||||| ||| ||| :|::|:||:|||||  |||||||  ||:|||:|
Db        234 GDCVEGFVEIFDMLLATASRFRMLKLKPEEFVCLKAIILLNSGAFSFCTGTMEPLHNSVE 293

Qy        426 LAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRH 467
              : ::|: :||||:  |::|| |:|||| | | ||:||||:||
Db        294 VHNMLDTITDALIHHISQSGCSAQQQSRRQAQLLLLLSHIRH 335


RESULT 22
Q91984
ID   Q91984       PRELIMINARY;       PRT;   276 AA.
AC   Q91984;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Estrogen receptor alpha (Fragment).
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Legler J., Broekhof J.L.M., Brouwer A., Lanser P.H., Murk A.J.,
RA   van der Saag P.T., Vethaak D.A., Wester P., Zivkovic D.,
RA   Van der Burg B.;
RT   "A novel in vivo bioassay for (xeno-)estrogens using transgenic
RT   zebrafish.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; A1275410; CAB77022.1; -.
DR   HSSP; P03372; 1ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   Pfam; PF00104; hormone_rec; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   SMART; SM00430; HOLI; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     276    276
SQ   SEQUENCE   276 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             24.2%;  Score 679;  DB 13;  Length 276;
  Best Local Similarity   51.9%;  Pred. No. 1.1e-54;
  Matches  147;  Conservative   47;  Mismatches   61;  Indels   28;  Gaps    6;

Qy        212 VGMVKGCGSRKERGGVRIVKKQR   - -----SADEQLHPAG---------KAKRSGGHA 252
              |||:| | |:::| | | |||:|         |: :|   ||              : |||||
Db          1 VGMMKGGIRKDR-GGRSVRPFRPRSSNFDRDKSSSDQCSKAGVRTIGPQDKPKKRSGGVV 59

Qy        253 PRVRELLLDALSPEQLVLTLLEAEPPHVLI-SRPSAPFTEASMMMSLIKLADKELVHMIS 311
              :       :|||:|::| || |||| |    : | |:|| :|| || :||||||||||:
Db         60 STL------CMSPDQVLTLLIGAEPPAVCSRQRSRPYTEITMMSLLINMADKELVHMIA 113

Qy        312 WAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVE 371
              |||:||| :||| |||:|||| |:|||:||:|||  ||||||| ||:||| ||:|||
Db        114 WAKKVPGFQDLSLHDQVQLLESSWLEVLMIGLIWRSIHSPGKLIFAQDLILDRSEGECVE 173

Qy        372 GILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAH-LL 430
              |: |||||||| :||| ||||: :|:::|:||:|||:|| :   :: :      :   :|
Db        174 GMAEIFDMLLATVARFRSLKLQLEEFVCLKAIILINSGAFSFCSSPVEPLMDNEMVQCML 233

Qy        431 NAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGM 473
              : :||||:: |:||| | | || | | ||:||||:|| ||||
Db        234 DNITDALIYCISKSCASLQLQSRRQAQLLLLLSHIRHMSNKGM 276


RESULT 23
Q90Z59
ID   Q90Z59      PRELIMINARY;      PRT;   200 AA.
AC   Q90Z59;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Estrogen receptor beta (Fragment).
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21235189; PubMed=11337474;
RA   Robinson-Rechavi M., Marchand O., Escriva H., Bardet P.L., Zelus D.,
RA   Hughes S., Laudet V.;
RT   "Euteleost Fish Genomes are Characterized by Expansion of Gene
RT   Families.";
RL   Genome Res. 11:781-788(2001).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF342950; AAK76402.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1      1
FT   NON_TER     200    200
SQ   SEQUENCE   200 AA;  22638 MW;  [illegible] CRC64;

  Query Match             21.2%;  Score 594.5;  DB 13;  Length 200;
  Best Local Similarity   59.5%;  Pred. No. 5.1e-47;
  Matches  119;  Conservative   22;  Mismatches   40;  Indels   19;  Gaps    8;

Qy        202 QACRLRKCYEVGMMKGGVRKERGGRPIQVRPQRSADEQLHPAGKAKRSGGHAPPVKELIT 261
              |||||||||||||: || ||||| ||  | :|:  : :| :|      | : :  |  |
Db          1 QACRLRKCYEVGMMNGGVRKERGGRPIQVRPQRSAGEQLHPAGKAKRSGGVVGLRGQSQQHIEPPLR 59

Qy        262 AL----------------SPEQLVLTLLEAEPPHVLISRP-SAPFTEASMMMSLTKLAD 303
              :|                |||||| :|||||| : :  |   |:||||||||||| |||
Db         60 SLSAPLSSGGRAEGRGLNYSPEQLVSCILEAEPPQIYLREPVKKPYTEASMMMSLTSLAD 119

Qy        304 KELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLD 363
              |||| |||||||||||||:| ||| ||| ||:::|:|||||:|||||| ||| |:
Db        120 KELVLMISWAKKIPGFVELILSDQVHLLECCWLDILMLGLMWRSVDHPGKLIFTPDLKLN 179

Qy        364 RDEGKCVEGILEIFDMLLAT 383
              |:|| |||||:||||| | |
Db        180 REEGNCVEGIMEIFDMFLTT 199


RESULT 24
Q9JM97
ID   Q9JM97      PRELIMINARY;      PRT;   139 AA.
AC   Q9JM97;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Estrogen receptor beta (Fragment).
GN   ESR2 OR ERB.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=21311378; PubMed=11418189;
RA   Ishibashi O., Kawashima H.;
RT   "Cloning and characterization of the functional promoter of mouse
RT   estrogen receptor beta gene.";
RL   Biochim. Biophys. Acta 1519:223-229(2001).
DR   EMBL; AB034957; BAA92313.1; -.
DR   MGD; MGI:109392; Esr2.
KW   Receptor.
FT   NON_TER     139    139
SQ   SEQUENCE   139 AA;  15215 MW;  4703615AE8593DAB CRC64;

  Query Match             19.4%;  Score 544;  DB 11;  Length 139;
  Best Local Similarity   84.2%;  Pred. No. 1.5e-42;
  Matches  101;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYPAMTFYSPAVMNYSIPS 60
              |:|||||||| ||:|||||||||||||| |||||||:| ||| |||||||||||||:||
Db         20 MEIKNSPSSLTSPASYNCSQSILPLEHGFIYIPSSYVESHHEYSAMTFYSPAVMNYSVPS 79

Qy         61 NVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
              :  |||||| ||| |||||||| |||||  | | | ||||||||||||||||||||||
Db         80 STGNLEGGPVRQTASPNVLWPTSGHLSPLATHCQSSLLYAEPQKSPWCEARSLEHTLPVN 139


RESULT 25
Q9VSE9
ID   Q9VSE9      PRELIMINARY;      PRT;   484 AA.
AC   Q9VSE9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01 MAY 2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG7404 protein (GH28308P) (Estrogen related receptor).
GN   ERR OR CG7404.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN BERKELEY;
RX   MEDLINE 20196006; PubMed 10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Lehman J., Wherry S., Talmadge C., Davis J.R., Kiss I., Endy J.D.,
RA   Chase B.A., Sumegi J.;
RT   "Identification and characterization of the Drosophila ortholog of
RT   vertebrate estrogen-related receptor gene and its product.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AE003556; AAF50473.1;  -.
DR   EMBL; AY051632; AAK94056.1;  -.
DR   EMBL; AF359420; AAL37553.1;  -.
DR   HSSP; P03372; 1HCQ.
DR   FlyBase; FBgn0035849; ERR.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   Prodom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc finger.
SQ   SEQUENCE   484 AA;  54245 MW;  252E3AD07C81A3C1 CRC64;

  Query Match             18.8%;  Score 527.5;  DB 5;  Length 484;
  Best Local Similarity   40.7%;  Pred. No. 2.7e-40;
  Matches  157;  Conservative   83;  Mismatches  179;  Indels   93;  Gaps   19;

Qy          36 YVDSHHEYPAMTFYSPAVMNYSIPSNVINLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQL 95
               ::    : :: : :|:        | |  : |     :|:|       ||   |||
Db           9 HIKQEVDIPSASCESPS-------SKSTATQSGINGLKSSPSV----  --SP  -ERQL 50

Qy          96 SHLYAEPQKSPWCEARSLEHTLPVNRETLKRK- ---- ------VSGNRC -ASPVTGP 140
                        |  |:  ::  :|  : ::||                ||     ||  |
Db          51 ----CSSTTSLSCDLHNV--SLSNDGDSLKGSGTSGGNGGGGGGGTSGGNATNASAGAGS 104

Qy         141 CSKRD--AHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRR 198
               || ||     | || | |||:|||| ||| |||||||:||: :| ||| |:| |:| ||
Db         105 GSVRDELRRLCIVCGDVASGFHYGVASCEACKAFFKRTIQGIEYTCPANNECEINKRRR 164

Qy         199 KSCQACRLRKCYEVGMVKCGSR--RERCGYRLVRRQ--RSADEQLHCAGKAKRSGGHAPR 254
               |:|||| :||  :||:| | |  | | |  :  ||    :: : :    ::  :     :
Db         165 KACQACRFQKCLLMGMLKEGVRLDRVRGGRQKYRRNPVSNSYQTMQLLYQSNTTSLCDVK 224

Qy         255 VRELLLDALSPEQLVLTLLEAEPPHV---LISRPSAPFTEASMMMS------------- 297
               : | :|::  |: |   ::  || |    |:  | : |: :
Db         225 ILE-VLNSYEPDAL---SVQTPTPQVHTTSTINDEASSSSGSIKLESSVVTPNGTCIPQN 280

Qy         298 ------------LTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
                          |: : ||||| :| |||:|||(::) | ||::||:  | |:| : | :
Db         281 NNNNDPNEILSVLSDIYDKELVSVIGWAKQIPGFIDLPLNDQMKLLQVSWAEILTLQLTF 340

Qy         346 RSIDHPGKLIFAPDLVLDRDEGK-CVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMI 404
               ||:   ||| || |: :|    | |  |  | :   :    |   :  : :||  :||::
Db         341 RSLPFNGKLCFATDVWMDEHLAKEC--GYTEFYYHCVQIAQRMERISPRREEYYLLKALL 398

Qy         405 LLNSSMYPLVTATQDADSSRKLAH-LLNAVIDALVWVIAKSGISSQQQSMRLANLLMLLS 463
               | |  :         |  | |     :||:: | :  :    | :| |||      ||:||
Db         399 LANCDI ----LLDDQSSLRAFRDTILNSLNDVVYLLRRSSAVSHQQQ------LLLLLP 447

Qy         464 HVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                :| | :         :    |: :  | ||||
Db         448 SLRQADDILRRFWRGIARDEVITMKKLFLEML 479


RESULT 26
Q8WS79
ID   Q8WS79      PRELIMINARY;      PRT;   496 AA.
AC   Q8WS79;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen-related receptor splice variant.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lehman J., Wherry S., Talmadge C., Davis J.R., Endy J.D., Chase B.A.,
RA   Kiss I., Sumegi J.;
RT   "Identification and characterization of the Drosophila ortholog of
RT   vertebrate estrogen-related receptor gene and its product.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF359421; AAL37554.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
```

```
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   496 AA;  54529 MW;  5D67E5AE8E9DCB32 CRC64;

  Query Match             18.8%;  Score 527.5;  DB 5;  Length 496;
  Best Local Similarity   30.7%;  Pred. No. 2.8e-40;
  Matches  157;  Conservative   83;  Mismatches  179;  Indels   93;  Gaps   19;

Qy          36 YVDSHHEYPAMTFYSPAVMNYSIPSNVINLEQGPQPQIISPNVIWPIPGHLSPLVVHPQL 95
               ::    : |: : :|||:        |  |   : |     :||:|         ||    |||
Db          21 HIKQEVDIPSASCFSPS-------SKSTATQSGTNGIKSSPSV--------SP---PPQL 62

Qy          96 SHLYAEPQKSPWCEARGLEHTLPVNPETLKRK           ----VSGNKC--ASPVIGP 140
                     |  |:  ::  :|  : ::||                  ||     ||   |
Db          63 ----CSSTTSLSCDLHNV--SLSNDGDSLKGSGTSGGNGGGGGGTSGGNATNASAGAGS 116

Qy         141 GSKRD--AHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRR 198
               || ||     | || | |||:|||| ||| |||||||:|||: :| ||| |:| |:| ||
Db         117 GSVRDEIRRLCIVCGDVASGFHYGVASCEACKAFFKRTIQGNIEYTCPANNECEINKRRR 176

Qy         199 KSCQACRLRKCYEVGMVKGGSR--RERGGYRLVPRQ--RSADPQLHGACKAKRSCCHAPR 254
               |:|||| :||  :||:| | |  | | | :  ||    :: : :    ::  :      :
Db         177 KACQACRFQKCLLMGMLKEGVRLDRVRGGRQKYRRNPVSNSYQTMQLLYQSNTTSLCDVK 236

Qy         255 VRFLLDALSPEQLVLTLLEAEPPHV---LISPPSAPFTEASMMMS-------------- 297
               : | :|::  |: |    ::  || |    |:   |  :  |: :
Db         237 ILE-VLNSYEPDAL---SVQTPPPQVHTTSITNDEASSSSGSIKLESSVVTPNGICIFQN 292

Qy         298 ------------LTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMW 345
                         |: : ||||| :| ||| :|||||:: | | ||::||:  | |:| : | :
Db         293 NNNNDPNEILSVLSDIYDKELVSVIGWAKQIPGFIDLPLNDQMKLLQVSWAEILTLQLIF 352

Qy         346 RSIDHPGKLIFAPDLVLDRDEGK-CVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMI 404
               ||:   ||| || |: :|    | |  |  | :   :    |   :  : :||  :|||::
Db         353 RSLPFNGKLCFATDVWMDEHLAKEC--GYTEFYYHCVQIAQRMERISPRREEYYLLKALL 410

Qy         405 LLNSSMYPLVTATQDADSSPRLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLS 464
               | |  :        |  | |     :||:: | :  :   | :| |||      ||:||
Db         411 LANCDI-----LLDDQSSLRAFRDTILNSLNDVVYLLRHSSAVSHQQQ---  --LLLLLP 459

Qy         464 HVRHASNKGMEHLLNMKCKNVVPVYDLLLEML 495
                :| | :        :    |: :  | ||||
Db         460 SLRQADDILRRFWRGIARDEVIFMKKLFLEML 491


RESULT 27
Q923V1
ID   Q923V1      PRELIMINARY;      PRT;   101 AA.
AC   Q923V1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Estrogen receptor beta (Fragment).
GN   ESR2.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang X.-H., Das S.K., Paria B.C.;
RT   "Mesocricetus auratus estrogen receptor beta (Esr2), mRNA sequence.";
RL   Submitted (JUN 2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF394639; AAK77002.1; -.
DR   InterPro; IPR001230; Prenyl_site.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1      1
FT   NON_TER     101    101
SQ   SEQUENCE   101 AA;  11328 MW;  13C22D39A8E029E0 CRC64;

  Query Match             18.7%;  Score 525;  DB 11;  Length 101;
  Best Local Similarity   92.1%;  Pred. No. 5.8e-41;
  Matches   93;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         107 WCEAPSLEHILPVNPETLKPKVSGNPCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSC 166
               | |||||||||| :|||||||:||  |||||| | ||||||||||||||||||||||||
Db           1 WYEARSLEHILPGSRETLKRKLSGGSCASPVISPSSKRDAHFCAVCSDYASGYHYGVWSC 60

Qy         167 PGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLR 207
               |||||||||||||||||||||||||||||||||||||||||
Db          61 PGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLR 101


RESULT 28
O46402
ID   O46402      PRELIMINARY;      PRT;   159 AA.
AC   O46402;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01 MAR 2002 (TrEMBLrel. 20, Last annotation update)
DE   Estrogen receptor alpha (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hou S., Winters T.A.;
RT   "Swine estrogen receptor alpha hormone-binding domain, exon V to
RT   VII.";
RL   Submitted (NOV 1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF035775; AAB88509.1; -.
DR   HSSP; P03372; 3ERD.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   Pfam; PF00104; hormone_rec; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   SMART; SM00430; HOLI; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     159    159
SQ   SEQUENCE   159 AA;  18454 MW;  83E44055?D8ECA98 CRC64;

  Query Match             18.4%;  Score 517.5;  DB 6;  Length 159;
  Best Local Similarity   59.5%;  Pred. No. 5.2e-40;
  Matches   94;  Conservative   39;  Mismatches   24;  Indels    1;  Gaps    1;

Qy         338 VLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEY 397
               :|:|:|:|||:|||||:|||:|:|||:|||||||:|||||||||:|||| : || :|:
Db           2 ILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEF 61

Qy         398 LCVKAMILLNSSMYPLVTAT-QDADSSPRLAHLLNAVTDALVWVIAKSGISSQQQSMRLA 456
               :|:|::|||||  :|  :::| :  :    :  :|: :|| |: ::||:|:: |||  |||
Db          62 VCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHRRLA 121

Qy         457 NLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEM 494
                ||::||| || |||||||| ||||||||:|||||||
Db         122 QLLLILSHIRHMSNEGMEHLYNMKCKNVVPLYDLLLEM 159


RESULT 29
```

```
Q95JC1
ID   Q95JC1       PRELIMINARY;       PRT;   249 AA.
AC   Q95JC1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ERalpha protein (Fragment).
GN   ERALPHA.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDIDYMIS;
RA   Kaeppler Hanno K.G.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDIDYMIS;
RA   Telgmann R.;
RT   "Epididymal epithelium immortalized by Simian virus 40 large T
RT   antigen: a model to study epididymal gene expression.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AJ313195; CAC42096.1; -.
DR   InterPro; IPR001292; Oestrgn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF02159; Oest_recep; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1       1
FT   NON_TER     249     249
SQ   SEQUENCE   249 AA;  27918 MW;  71C23C6540164692 CRC64;

  Query Match             18.0%;  Score 506;  DB 6;  Length 249;
  Best Local Similarity   57.1%;  Pred. No. 1.1e-48;
  Matches  100;  Conservative   23;  Mismatches   30;  Indels   22;  Gaps    4;

Qy        142 SKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSC 201
              | ::  :|||||:|||||||||||||||||||||||||||||:|||||||||||||||||
Db         79 SAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSC 138

Qy        202 QACRLRKCYEVGMVEGGSKRFKRGRLAKNKKLLDENLVPRQRSAIEQLHCAGKAKRSGGHAPRVR- --  256
              |||||||||||||||:| | |::| | |::: :|  |:      |: :   | :  ||
Db        139 QACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDD- --EEGRNEVGSSGDVRTSSLW 193

Qy        257  ELLLD -    - -- ALSPEQLVLTLLEAEPPHVLIS-RPSAPFTEASMMMSLT 299
                ||:          :|: :|:|  |||||||  :     || ||:|||||  ||
Db        194 PSPLLIKHTKKNSPALSLTADQMVSALLEAEPPIIYSDYDPSRPFSEASMMGLLT 248


RESULT 30
Q9IBD5
ID   Q9IBD5       PRELIMINARY;       PRT;   710 AA.
AC   Q9IBD5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Progesterone receptor.
GN   PR.
OS   Anguilla japonica (Japanese eel).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC   Anguilla.
OX   NCBI_TaxID=7937;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=20088790; PubMed=10620698;
RA   Todo T., Ikeuchi T., Kobayashi T., Kajiura-Kobayashi H., Suzuki K.,
RA   Yoshikuni M., Yamauchi K., Nagahama Y.;
RT   "Characterization of a testicular 17,20-dihydroxy-4-pregnen-3-one (a
RT   spermiation-inducing steroid in fish) receptor from a teleost,
RT   Japanese eel (Anguilla japonica).";
RL   FEBS Lett. 465:12-17(2000).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AB032075; BAA89549.1; -.
DR   HSSP; P06536; 1RGD.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   710 AA;  78819 MW;  0C8149B05CA055AA CRC64;

  Query Match             18.0%;  Score 504;  DB 14;  Length 710;
  Best Local Similarity   31.4%;  Pred. No. 6.8e-48;
  Matches  135;  Conservative   84;  Mismatches  145;  Indels   66;  Gaps   19;

Qy         86 LSPLVVHR -- QLSHLYAEPQKSP-WCEAR- SLEHTLPVN-----RETLKRKVSGN 134
              :  | ||:    :||   | |  || ||::   | :| | ::          |   | |
Db        271 IQTLPVHKSEPERLSASSA-PADSPEWCQSIGPSEDHHILQIDYLSPAGLHSTC--KYSST 327

Qy        135 RCASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQC 194
                 |   |   :|    | :| | ||| |||| :|  || ||||:::|||:|:|   | |
Db        328 NAYSSYLGVLPQR---VCVICGDEASGCHYGVLTCGSCKVFFKRAVEGHHNYLCAGRNDC 384

Qy        195 TIDKNRRKSCQACRLRKCYFVGMVKGGSRREPCG--- ---------YRLVKRQRSAIEQ 241
              :|| |||:| ||||||||: ||:  | : :: |             : |  |: | | |
Db        385 IVDKIRRKNCPACRLRKCYQAGMTLGGRKLKKLGALKAAGLIQALVAHSLTPRRLSGDSQ 444

Qy        242 LHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLIS--RPSAPFTEASMMM 299
                |    |  | |||     |||:  :::::||  | |: |    | |      ::
Db        445 --- --ALMPLGCLPGVREL---HLSPQ-  IISVLESIEPEVVYSGYDNSQPDMPNMLLN 493

Qy        300 SLTKLADKELVHMISWAKRIPGFVELSLEDQVRLLESCWMEVLMMGLMWRSIDH- PGKL 358
              || :| :::|: :: |:| :|||  | : ||: |::  || :::  | |||  :     |
Db        494 SLNRLGERQLLRIVKWSKSLPGFRSLHINDQMALIQYSWMSLMVFSLGWRSFQNVNSDYL 553

Qy        359 IFAPDLVLDRDEGKCVEGILEIFDMLLA---TTSRFRELKLQHKEYLCVKAMILLNSSMY 415
              |||||:|  :  :       |||: :|      |  |::  :|:||:| ::|||:
Db        554 YFAPDLILNEEYMR - -RSPIFDLCMAMQFIPQEFANLQVTKEEFLCMKVLLLLNTVPL 609

Qy        416 PLVIATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNK 475
              : :    |  |:  : :: :| |:   : ::|: :  |  |  :|  |: |:     |
Db        610 EGLKSQPQFDEMRQ--NYIHELTKAI--HLRENGVVACSQ--RFYHLTKLMDHMHDIVKK 663

Qy        476 GMEHLLNMKC 485
                |:: |
Db        664   ---LHLYC 668


RESULT 31
Q9DEV4
ID   Q9DEV4       PRELIMINARY;       PRT;   583 AA.
AC   Q9DEV4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Progesterone receptor PR.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20524048; PubMed=11050156;
RA   Bayaa M., Booth R.A., Sheng Y., Liu X.J.;
RT   "The classical progesterone receptor mediates Xenopus oocyte
RT   maturation through a nongenomic mechanism.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12607-12612(2000).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AY007198; AAG01366.1; -.
DR   HSSP; P06536; 1RGD.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR000128; Progest_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02161; Prog_receptor; 2.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   583 AA;  66156 MW;  638D3572C32BD86C CRC64;

  Query Match             17.5%;  Score 491;  DB 13;  Length 583;
  Best Local Similarity   25.9%;  Pred. No. 8.5e-37;
  Matches  143;  Conservative   97;  Mismatches  162;  Indels  146;  Gaps   23;

Qy          2 DIKNSPSSLNSPSS-------------YNCSQSILPLEHGS------IYIPSSYVDSHHE 42
              |  :|| : ::| :              |:|       : ||:         :|||:
Db        106 DYGSSPRNPSTPDTDPSLDFILYKNEDYDC----FKISHGNTNEDSGCILPST------ 154

Qy         43 YPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHL-YAE 101
                | | | |  :|        :  | ::|  | :              || :: |
Db        155 -SAQTIYQPLSLN---GHQYVTFQPTPMKETYLPQI--------------QLPYVTYIR 195

Qy        102 PQKSPWCFARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHY 161
                  |       |  :| : | | :|:                 | :| | ||| ||
Db        196 SDGDP-------ERGIPFSFEMLPQKI-------------------CLICGDEASGCHY 228

Qy        162 GVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRR 221
              || :|  || |||||:|:|| :|:|   | | :|| |||:| :|||||| : |||  | :
Db        229 GVLTCGSCKVFFKRAIEGHQNYLCAGRNDCIVDKIRRKNCPSCRLRKCCQAGMVLGGRKF 288

Qy        222 EPCGYRLVRRQRSADE------QLHCAGKAKR--SGGHAPRVPELLLLDALSPEQLVL 270
              :: |   ::  | |:        | |     |  |  |  ::       :|| :|
Db        289 KKFG--RIKTGREIDTVVLQSPPTISLFCQQIILPRISNSSAQEIQ------FTPP--LI 338

Qy        271 TLLEAEPPHVLIS-RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQV 328
              :|::  | |: :    : | | :::: || :| ::|| :: |:| :|||  | : ||:
Db        339 QILQSIEFEVVYAGYDTTQPPTPSAILSSLNQLGERQLVSVVKWSKSLPGFRNLHIDDQI 398

Qy        329 RLLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGK-------CVEGILEIFDM 379
              ||:  || :::  | |||  |     | ||||||:|:    |       |:
Db        399 TLLQYSWMSLMVFALGWRSYQHVSGQMLYFAPDLILNEQRMKDSSFYTLCLS-------  450

Qy        380 LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRK--LAHLLNAVTDAL 437
              :      | :|:: |:|:|||:|||:|||:     : :  : |  |   :  |::
Db        451 MWQLPQEFMKLQVTHFEFLCMKALLLFNTIPLEGLKSQTNFDEMRSNYIRELAKAIS---  507

Qy        438 VWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHL   LNMKGKNV-  -      Vr 486
              :  |  |:| |:  :|  |:                |  |   |:: | |          |
Db        508 --LRHKGVIASSQRFYQLTKLM  - --------DSMHFLVKQLHLYCLNTFLQSRSLSVE 554

Qy        487 VYDLLLEMLNAHV 499
              ::: |:::| :
Db        555 FPEMMSEVISAQL 567


RESULT 42
Q8UVY3
ID   Q8UVY3      PRELIMINARY;      PRT;   703 AA.
AC   Q8UVY3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Oviduct progesterone receptor (Fragment).
GN   PR-2.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tian J., Ruderman J.V.;
RT   "Cloning of a progesterone receptor xPR-2 from Xenopus laevis
RT   oviduct.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF319596; AAL55260.1; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR000128; Progest_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02161; Prog_receptor; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   Receptor.
FT   NON_TER       1       1
SQ   SEQUENCE   703 AA;  79046 MW;  942570D87854D447 CRC64;

  Query Match             17.4%;  Score 489;  DB 13;  Length 703;
  Best Local Similarity   26.2%;  Pred. No. 1.7e-36;
  Matches  145;  Conservative   98;  Mismatches  164;  Indels  146;  Gaps   24;

Qy          2 DIKNSPSSLNSPSS-------------YNCSQSILPLEHGS------IYIPSSYVDSHHE 42
              |  :|| : ::| :              |:|       : ||:         :|||:
Db        226 DYGSSPLNPSTPDTDPSLDFILYKNEDYDC----FKISHGNTSEDSGCILPST------- 274

Qy         43 YPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHL-YAE 101
                | | | |  :|        :  | ::|  | :              || :: |
Db        275 -SAQTIYQPLSLNGH---QHVI-FQPTPMKETYLPQI--------------QLPYVTYIR 315

Qy        102 PQKSPWCFARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHY 161
                  |       |  :| : | | :|:                 | :| | ||| ||
Db        316 SDGDP-------ERGIPFSFEMLPQKI-------------------CLICGDEASGCHY 348

Qy        162 GVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRR 221
              || :|  || |||||:|:|| :|:|   | | :|| |||:| :|||||| : |||  | :
Db        349 GVLTCGSCKVFFKRAIEGHQNYLCAGRNDCIVDKIRRKNCPSCRLRKCCQAGMVLGGRKF 408

Qy        222 EPCGYRLVRRQRSADE---------QLHCAGKAKR--SGGHAPRVPELLLLDALSPEQLVL 270
              :: |   ::  | |:         | |     |  |  |  ::       :|| :|
Db        409 KKFG--RIKTGREIDIVVLQSPPTLSLECQQIILRRISNSSAQEIQ------FTPE--LL 458

Qy        271 TLLEAEPPHVLIS-EPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQV 328
              :|::  | |: :    : | | :::: || :| ::|| :: |:| :|||  | : ||:
Db        459 QILQSIEFEVVYAGYDNIQPEIPSALLSSLNQLGERQLVSVVKWSKSLPGFRNLHIDDQI 518

Qy        329 RLLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGK-------CVEGILEIFDM 379
              ||:  || :::  | |||  |     | ||||||:|:    |       |:
Db        519 TLLQYSWMSLMVFALGWRSYQHVSGQMLYFAPDLILNEQRMKDSSFYTLCLS-------- 570
```

```
Qy          380 LLATTSRFRELKLQHKEYICVKAMILLNSSMYPLVTATQDADSSRK--LAHLLNAVTDAL 437
                :       | :|:: |:|:||:|::|||:      : :  : | :|   :  | |::
Db          571 MWQLPQEFMKLQVTHEEFLCMKALLLLNTIPLEGLKSQTNFDEMRSNYIRELAKAIS-   627

Qy          438 VWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHL--LNMKCKNV---    VP 486
                  :  | :|  |:  :|  |:                 | |  |:: | |       |
Db          628   LRHKGVIASSQRFYQLTKLM   - -  - --DSMHELVKQLHLYCLNTFLQSRSLSVE 674

Qy          487 VYDLLLEMLNAHV 499
                ::: |:::| :
Db          675 FPEMMSEVISAQL 687


RESULT 44
Q96102
ID   Q96102       PRELIMINARY;       PRT;   423 AA.
AC   Q96102;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Unknown (Protein for IMAGE:4301880) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UTERUS;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; BC007915; AAH07915.1;  -.
DR   InterPro; IPR000515; BPD_transp.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00402; BPD_TRANSP_1NN_MEMBR; UNKNOWN_1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc finger.
FT   NON_TER       1       1
SQ   SEQUENCE   423 AA;  45341 MW;  75D5BA8595802SDC CRC64;

  Query Match             17.0%;  Score 475.5;  DB 4;  Length 423;
  Best Local Similarity   47.5%;  Pred. No. 1.1e-45;
  Matches  124;  Conservative   56;  Mismatches  110;  Indels   21;  Gaps    8;

Qy          174 KRSTQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRREG--GYRLVPRQ 232
                ||:|||  :| |||:|:| | | |||:|||||  ||  |||:| | | :|  | |   ::
Db            4 KRETQGSTEYSCPASNEGEIKKRRKRACQACRFTKCLRVGMLKEGVRLDRVRGGRQKYKR 63

Qy          233 RSADEQL    - HCAGKAKRSGGHAPPVRELLLDALSPEQLVLTLLEAEPPHV-LISRPS 286
                |  : |       ||   :|| ||     :||     || || ||  |  :   |:
Db           64 RPEVDPLGFPGEFFAGPGAVAGG--PRKTAAPVNA   - --LVSHLLVVEPEKLYAMPDPA 116

Qy          287 APFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWR 346
                 |      : :|  | |:|:|  ||||| |||  ||| ||: :|:| ||||||:|:  |
Db          117 GPDGHLPAVATLCDLFDREIVVTISWAKSIPGFSSLSLSDQMSVLQSVWMEVLVLGVAQR 176

Qy          347 SIDGHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATSSRFRELKLQHKEYICVKAMILL 406
                |:    :| || ||||| :||    |: |:   ||    | : |:|: :||: :||: |
Db          177 SLPLQDELAFAEDLVLD-EEGARAAGLGELGAALLQLVRRLQALRLEREEYVLLKALALA 235

Qy          407 NSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVR 466
                ||    :    :|||:  :|   |:     | :   ::|     :  |   ||: |  :|
Db          236 NSDSVHI     EDAEAVEQLREALHEA   LLEYEAGRAGPGGGAERRAGRLLLTLPLLR 289

Qy          467 HASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
                  : | : |   :| :  ||:: | |||| |
Db          290 QTAGKVLAHFYGVKLEGKVPMHKLFLEMLEA 320


RESULT 44
Q9DDD9
ID   Q9DDD9       PRELIMINARY;       PRT;   732 AA.
AC   Q9DDD9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Progesterone receptor.
GN   XPR-1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 20570484; PubMed 11114187;
RA   Tian J., Kim S., Heilig E., Ruderman J.V.;
RT   "Identification of XPR-1, a progesterone receptor required for Xenopus
RT   oocyte activation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:14358-14363(2000).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF279335; AAG42362.1; -.
DR   HSSP; P06536; 1RGD.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR000128; Progest_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02161; Prog_receptor; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   732 AA;  82022 MW;  BF0BA3A14F022199 CRC64;

  Query Match             16.7%;  Score 468.5;  DB 13;  Length 732;
  Best Local Similarity   26.0%;  Pred. No. 1.4e-44;
  Matches  143;  Conservative   97;  Mismatches  173;  Indels  147;  Gaps   24;

Qy            2 DTKNSPSSLNSES--SYN-------SQSITPLPHGS-----YYTPSSYVDSHHEY 43
                |  :|| : ::|    | |         |     : :|:          :|||:
Db          252 DYGSSPQNPSTPDPDPSLNEILYKNDSACDYDSFKISYGNTSEGSGCVLPST-       303

Qy           44 PAMTFYSPAVMNYSTPSNVINLEGGPGRQTISPNVLWPTPGHLSPLVVHRQLSHL YAEP 102
                | | | |  :|            |   | |  :  | :|| :    || : :|
Db          304 SAQTIYQPLSLN--  ---------GRQHITFQPTSMKET--YLSEI  -  -QLPYVSYIRS 345

Qy          103 QKSPWCEAKSLEHTLPVNKEILKKKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYG 162
                  :|       :   | : ||  :|:                      | :| | ||| |||
Db          346 DGNP       DPGSPFSFETFPQKI                    CLICGDEASGCHYG 378

Qy          163 VWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRRE 222
                | :|  || ||||:|:| ::|:|    | ) :|| |||:| :|||||| : ||| |  : :
Db          379 VLICSSCNVEFKKAIEGQHNYLCAGRNDCIVDKIRRKNCPSCRLRKCCQAGMVLGGRKFK 438

Qy          223 RGGYRLVPRQRSADE------  QLHCAGKAKR-SGGHAPRVRELLLDALSPEQLVLT 271
                : |   ::  |  |           | |     |  |   | ::       :||  :|
Db          439 REG--RIKIGREIDIVVLQSPPTLSLEQQQIILIRRISNSSAQEIQ -  - FTPE--LLQ 488

Qy          272 LLEAEPPHVLIS--RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVR 329
                :|::  | |: :    : | | ::: || :| ::||| :: |:| :||| | : ||:
Db          489 ILQSIEPEVVYAGYDNTQPETPSALLSSLNQLCERQLVCVVKWSKSLPGFRNLHIDDQIT 548

Qy          330 LLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGK ----- CVEGILEIFDML 380
                ||:  || :::  | |||  |     | ||||:|:    |      |:       :
```

```
Db        549 LLQYSWMSLMVFALGWRSYQHVSGQMLYFAPDLILNEQRMKDSSFYTLCLS--------M 600

Qy        481 LATTSRFREIKIQHKEYICVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWV 440
              | :|::: |:|:|:|:|:::|||:   ||       :      :: :  : | : :
Db        601 WQIPQFFMKTQVTHFFFLCMKAIILINT--IPLEGLRSQTNFEEMRSNYIRELAKA-IGL 657

Qy        441 IAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHL---LNMKCKNV--------VPVYD 489
              |  |:| |:  :|  |:                |  |   |:: | |         |  :
Db        658 RHKGVIASSQRFYQLTKLM----------DSMHELVKQLHLYCLNTFLQSRSLSVEFPE 709

Qy        490 LLLEMLNAHV 499
              :: |:::| :
Db        707 MMSEVISAQL 716


RESULT 35
Q8QGX5
ID   Q8QGX5      PRELIMINARY;      PRT;   689 AA.
AC   Q8QGX5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Progestogen receptor type 2
OS   Anguilla japonica (Japanese eel).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC   Anguilla.
OX   NCBI_TaxID=7937;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=21629085; PubMed=11755535;
RA   Ikeuchi T., Todo T., Kobayashi T., Nagahama Y.;
RT   "A novel progestogen receptor subtype in the Japanese eel, Anguilla
RT   japonica.";
RL   FEBS Lett. 510:77-82(2002).
DR   EMBL; AB028024; BAB85993.1; -.
KW   Receptor.
SQ   SEQUENCE   689 AA;  76316 MW;  F061DBD5AC7A72A6 CRC64;

  Query Match             16.7%;  Score 467.5;  DB 13;  Length 689;
  Best Local Similarity   27.3%;  Pred. No. 1.6e-34;
  Matches  148;  Conservative   83;  Mismatches  208;  Indels  103;  Gaps   20;

Qy          4 KNSPSSLNSPSSYNC------SQSILPLEHGSIYI-PSSYVD--------------SHHE 42
              | | | ::| |  ||       |   :|  | |  ||  :               :||
Db        145 KASLSGVSSSSVGNCKFIKDFQDSSSSMFPQSPYFHPSGNITTSNSSYGTCEEDSATHHP 204

Qy         43 YPAMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTP------------GHLSPIL 89
                | |:       ||            : |     :| |                | ||
Db        205 PHMFTDYNRTTALPLIP-----------EITEDQFSFPYPVGEVVANSCLIGYGQRSPQ 252

Qy         90 VVHR------QLSHLYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNRCA--------S 135
                |      :||   . |.   ||..  |           |      | |           |
Db        253 NSIPFKSPLCKLSLPTSSPESQSWCQSTGLSEDQHFETCYLPPCEIRNICETHNSLKSHS 312

Qy        136 PVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQCHNDYICPATNQCTIDK 195
              |  |::   || :| | ||| |||| :|  || |:||::::|| :|:|    | | :||
Db        313 VYMGMLSQK---FCLICGDEASGCHYGVLTCGSCKVFYKRAVEGHQNYLCAGRNDCIVDK 369

Qy        196 NRRKSCQACRLRKCYEVGMVKCGSRREPG---------YRLVPRQRSADEQLHCAGKAK 246
               |||:| |||||||: ||   | : ::           ||      |  :     |:|
Db        370 IRRKNCPACRLRKCYQAGMTLGGRKMKKLSALKVLGLTQSLAVRSPLGASYE----GQAL 425

Qy        247 RSGGHAPPVPFLLLDAISPEQLVITLLEAPPPHVLISRPSA--PFTEASMMMSLTKLADK 304
              :   | ||||     :|:  :|:::||   |  : |   |  | |    :: ||  | ::
Db        426 ATLPSMPMVREL---QFTPQ--ILSILENIEPETVYSGYDATQPETPHLLLNSLNGLCER 480

Qy        305 ELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPGK--LIFAPDLVL 362
              :|: :: |:| :|||  | : || |:  ||  :::  | |||  :  :  | ||||||:|
Db        481 QLLWIVRWSKSLPGFRSLHINDQMTLIQYSWMSLMVFSLGWRSFQNVTREFLYFAPDLIL 540

Qy        363 DRDEGKCVEGILEIFDMLLA---TTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQD 419
                                                                        :: :       | |: :|       |  |:::  :|:|||:| ::|||:   ||
Db        541 GEEKMR----NSPISDLCMAMQIIPQAFDNLQVTKEEFLCMKVLLLLNT--VPLEGLRSQ 594

Qy        420 ADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNM 479
              |          :  :| |:   : : |: :  |  | :|  |:  :     |     :||:
Db        595 AQFDEMRHGYIRELTKAI--QLIERGVMASSQ--RFYHLTKLMDAMHEIVRK --  -VNL 645

Qy        480 KC 481
              |
Db        646 YC 647


RESULT 36
Q90ZM7
ID   Q90ZM7      PRELIMINARY;      PRT;   438 AA.
AC   Q90ZM7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Corticoid receptor (Fragment).
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21244662; PubMed=11331759;
RA   Thornton J.W.;
RT   "Evolution of vertebrate steroid receptors from an ancestral estrogen
RT   receptor by ligand exploitation and serial genome expansions.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:5671-5676(2001).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AY028457; AAK20940.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   Prodom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1       1
SQ   SEQUENCE   438 AA;  48935 MW;  BB6FEE6637BFE471 CRC64;

  Query Match             16.6%;  Score 466.5;  DB 13;  Length 438;
  Best Local Similarity   27.5%;  Pred. No. 1.1e-34;
  Matches  134;  Conservative   76;  Mismatches  166;  Indels  111;  Gaps   15;

Qy         47 TFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVL----WPTPGHLSPLVVHRQLSHLYAEP 102
              | : |:|   |  | ::|   |      ||| :    :| ||  ||       :    |
Db         13 TSFRPSVAT-SSASGISNFSNGNNFGFLSPNGVQQDGFPYPGFTSP-------AQSSVPP 64

Qy        103 QKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYHYG 162
              ||:                                          | :||| ||| |||
Db         65 QKA-----------------------------------------CLICSDEASGCHYG 81

Qy        163 VWSCEGCKAFFKRSIQCHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRE 222
              | :|  || |||||:::| ::|:|   | | ||| ||||| |||||||  : ||     : :
Db         82 VLTCGSCKVFFKRAVEGQHNYLCAGRNDCIIDKIRRKNCPACRLRKCIQAGMTLGARKLK 141

Qy        223 RCGYRLVRRQRSADEQLHCAGKAKRSGGHAPPVPFLLLDAISP--------FQLVITLLE 274
              : |       |||         |       |       :   ||              :: :|:
Db        142 KQGRVKGENQRSPASSTATTSSATPQ----PSSNSTAVTTFSPPPTGEPIFSPTLIAILQ 197

Qy        275 AFPPHVLIS--RPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLE 332
              |  | |::|    :  | | (: || :| ||:|| :: ||| :|||  | : ||: |::
Db        198 AIEPEVVMSGYDNTRSQTTAYMLSSLNRLCDKQLVSIVKWAKSLPGFRNLHIDDQMVLIQ 257

Qy        333 SCWMEVLMMGLMWRSIDHPGK--LIFAPDLVLDRDEGK-------CVEGILEIFDMLIAT 383
              || ::   : |||  |      | ||||||  |    :        |||          :
Db        258 YSWMGLMSFAMSWRSFQHTNSKLLYFAPDLVFDETRMQQSAMYQLCVE--------MRQV 309
```

```
Qy        384 TSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSP-- ---KLAHL--LNAVTD 435
              :  | :|::  :|:|||:||::|:        | |:   |:       :::::  ||
Db        310 SEDFMKLQVTSEEFLCMKAILLLS---  -  IVPQEGLKSQQCFEEMRTSYIRELNK -  359

Qy        436 ALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKN- -VVPVYDLLL 492
                 ||::  :: |  |  |  || ::  :| :|       :     |  |:::
Db        360      TIARTEKNAVQCWQRFYQLTKLLXCMQDLVSKLLEFCFAIFTQTQVWSVEFPDMMA 415

Qy        493 EMLNAHV 499
              |:::| :
Db        416 EIISAQL 422


RESULT 47
Q95L53
ID   Q95L53       PRELIMINARY;       PRT;   525 AA.
AC   Q95L53;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Retinoid X receptor beta (Fragment).
OS   Mustela vison (American mink).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
OC   Mustela.
OX   NCBI_TaxID=9667;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hua X., La P.;
RT   "Cloning and characterization of mink RXRB.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF418003; AAL14642.1; -.
DR   InterPro; IPR000536; Hormone_rec_liq.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc finger.
FT   NON_TER       1       1
SQ   SEQUENCE   525 AA;  59975 MW;  51ABB1875134582 CRC64;

  Query Match             16.4%;  Score 458.5;  DB 6;  Length 525;
  Best Local Similarity   30.2%;  Pred. No. 7.6e-34;
  Matches  140;  Conservative   63;  Mismatches  163;  Indels   97;  Gaps   17;

Qy         62 VTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN- 120
              :::  | ||    :|      ||              :: | ||   :     :||
Db        128 ISSSMGSPGLPPPAP--   - PG---------- ---FSGPVSSPQINSTV---SLPGGG 164

Qy        121 -  RETLKRKVSGNR - CASPVTGPGSKRDAHFCAVCSDYASGYHYGVWSCEGCKAFFK 174
               | :|  | | |   |  |  ||| :    ||:| | :|| |||||||||| |||
Db        165 SGPPFDVKPPVLGVRGIHCPPPPGGPGAGK--RICAICGDRSSGKHYGVYSCEGCKGFFK 222

Qy        175 RSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQRS 234
              |:|    | |      ||:|| :|  || || :||   || :   : |         |||
Db        223 RTIRKDLTYSCRDNKDCTVDKRQRNRCQYCRYQKCLATGMKREAVQEE-----  ---RQRG 274

Qy        235 ADEQLHCAGKAKRSGCHAPRVRELLLDALSPEQL-VLTLLEAEPPHVLI- -  -  -  - 282
               |:    |  : :||              :|||:: |  :||||   : :
Db        275 KDKD   GDGEGAGG-    -  -  -  --APEEMPVDRILEAE- --LAVEQKSDQGVEGP 314

Qy        283 -      --SRPSAPETEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW 335
                   | |: | |       :: : |||:|  :: |||:|| |  | | ||| || : :|
Db        315 GGTGGSGSSPNDPVT     - NICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGW 368

Qy        336 MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQH 394
               |:|:     ||||    :: | |  | : |:       |:  ||| :|    |: |::::
Db        369 NELLIASFSHRSIDVRDGILLATGLHVHRNSAHSA-GVGAIFDRVLTELVSKMRDMRMDK 427
```

```
Qy        395 KEYLCVKAMILLNSSMYPLVTATQDADSSPELAHLLNAVTDALVWVIAKSGISSQQQSMR 454
              |  |::|:|| |     |        |:     :|     | :    |     ||   |
Db        428 IELGCLRAIILFNPDAKGL  -  -  --SNFSEVEVLREKVYASLEAYCKQKYPEQQG  R 477

Qy        455 LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNA 497
              | ||: |  :|    | :|||    |      |:   |:||| |
Db        478 FAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA 520


RESULT 48
Q8I5C6
ID   Q8I5C6       PRELIMINARY;       PRT;   436 AA.
AC   Q8I5C6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   RXR-like protein.
GN   RXR.
OS   Biomphalaria glabrata (Bloodfluke planorb).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
OC   Planorbidae; Biomphalaria.
OX   NCBI_TaxID=6526;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bouton D., Mendonca R.L., Noel C., Escriva H., Cornette J., Laudet V.,
RA   Pierce R.J.;
RT   "A conserved RXR family nuclear receptor from the mollusc Biomphalaria
RT   glabrata.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY048664; AAL86461.1; -.
SQ   SEQUENCE   436 AA;  48225 MW;  941DE841B5CEBD4F9 CRC64;

  Query Match             16.3%;  Score 458;  DB 5;  Length 436;
  Best Local Similarity   28.5%;  Pred. No. 6.6e-34;
  Matches  144;  Conservative   68;  Mismatches  173;  Indels  120;  Gaps   16;

Qy          2 DIKNSPSSLNSPSSYNCSQSTLPLEHGSTYTPSSYVDSHHEYPAMTEYSPAVMNYSTPSN 61
              |||   ||| ||:|                        :| |     : |  |    ||
Db         34 DIKPDISSLTSPTS - - -  -  - - - -  - - - -   THGYYG- - - FGPGGM-   PSM 62

Qy         62 VTNLEGGPG RQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQKSPWCEARSLEHTLPVN 120
               :: :  || :|  || :  ||     || ::
Db         63 ASSTQPSPGPQQMHSPGMHSPTSSSMGSPPML- - - -  -  -  - - - - - - - - - - 93

Qy        121 RETLKRKVSGNRCASPVTGP---  ----GSKRDAHFCAVCSDYASGYHYGVWSCEGCKAF 172
                          | || :||          |    | ||:| | ||| ||||:|||||| |
Db         94 ------------CLSP-SGPSPSPGLPHSSLHTKHICAICGDRASGKHYGVYSCEGCKGF 140

Qy        173 FKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRRERCGYRLVRRQ 232
              |||:::    | |     | ||| :|  || ||  ||  :|| :   : ||      | :
Db        141 FKRTVRKDLIYACRDDKNCMIDKRQRNRCQYCRYMKCLSMGMKREAVQEER-----QRVK 195

Qy        233 RSADEQLHCAGKAKRSGCHAPRVRELLLDA-LSPEQLVLTLLEAEPPHVLISRPSAPFTE 291
               | ::          || :    | :|:| |: :  : | ::|:   |
Db        196 EKGDGEVE-----STSGANNDMPVEQILEAELAVDPKIDTYIDAQKDPV---------- 239

Qy        292 ASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWRSIDHP 351
              :: : |||:|  :: |||:|| | || | | ||| || : | |::|  |   |||
Db        240 ----INICQAADKQLFTLVEWAKRIPHFTELPLEDQVILLRAGWNELLIAGFSHRSIMAK 295

Qy        352 GKLIFAPDLVLDRDEGKCVEGILEIFDMLLA-TTSRFRELKLQHKEYLCVKAMILLNSSM 410
              :: |  | : |        |:  ||| :|    :: |::|:   |  |::|::| |
Db        296 DGILLATGLHVHRSSAH-QAGVGTIFDRVLTELVAKMRDMKMDKTELGCLRAVVLFNPDA 354

Qy        411 YPLVTATQDADSSPRLAHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASN 470
              | || |: :  |:  :        | :    ||     ::  | | || :  :|
Db        355 KGL-TAVQEVEQLREKVY-- ----ASLEEYTKS -RYPEEPGRFAKLLLRLPALRSIGL 404

Qy        471 KGMEHLLNMKCKNVVPVYDLLLEML 495
              | :|||   |     |:   |:|||
Db        405 KCLEHLFFFKLIGDQPIDTFLMEML 429
```

```
RESULT 39
Q9GLW0
ID   Q9GLW0      PRELIMINARY;      PRT:   939 AA.
AC   Q9GLW0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Progesterone receptor.
GN   PR.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lantinga-van Leeuwen I.S., van Garderen E., Mol J.A.;
RT   "Molecular cloning and cellular localization of the canine
RT   progesterone receptor.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF177470; AAG09282.1; -.
DR   HSSP; P06401; 1A28.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR000128; Progest_receptor.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02161; Prog_receptor; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   939 AA;  98417 MW;  14AB8E535A64F239 CRC64;

  Query Match             16.3%;  Score 458;  DB 6;  Length 939;
  Best Local Similarity   32.5%;  Pred. No. 1.8e-33;
  Matches  106;  Conservative   64;  Mismatches  126;  Indels   30;  Gaps   10;

Qy        149 CAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRK 208
              | :| | ||| |||| :|  || |||:::| ::|:|   | |  :|| |||:| |||||
Db        573 CLICGDEASGCHYGVITCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRK 632

Qy        209 CYEVGMVKGGSRKERGGYRLVRRQRSADEQL--HCAGKAKRSGCHAPPVRELLLDALSPE 266
              | : |||  |          ||  |. |        |    |      :|          :|
Db        633 CCQAGMVLGGRKFKK--FNKVRVVRTLDAVALPQPVGIPNESQALSQRI------SFSPS 684

Qy        267 Q-------LVLTLLEAFPPHVLISRPSA-PFTEASMMMSLTKLADKELVHMISWAKKIPG 318
              |       |:  |:  ||  :     :  | | ·|·· || ·| ···|· ·· |·| ·||
Db        685 QDIQLIPPLINLLMSIEPDVIYAGHDNTKPDTSSSLLTSLNQLGERQLLSVVKWSKSLPG 744

Qy        319 FVELSLFDQVRLLESCWMEVLMMGLMWRSIDHPG--KLIFAPDLVLDRDEGKCVEGILEI 376
              |  | : ||: |::  || ::: || |||  |     | ||||·|·  |       ·
Db        745 FRNLHIDDQITLIQYSWMSLMVFGLGWRSYKHVSGQMLYFAPDLILNEQRMK-ESSFYSL 803

Qy        377 FDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDA 436
              :    | :|::  ·|·||·| ··|||·   ||    ·   |  ·    ·  ·  ·
Db        804 CLTMWQIPQEFVKLQVSQEEFLCMKVLLLLNT--IPL----EGLRSQNQFEEMRSSYIRE 857

Qy        437 LVWVIA---KSGISSQQQSMRLANLL 459
              |:  |    |  :|| |:  :|  ||
Db        858 LIKAIGLRQKGVVSSSQRFYQLTKLL 883


RESULT 40
P70048
ID   P70048      PRELIMINARY;      PRT:   790 AA.
AC   P70048;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Androgen receptor alpha isoform.
GN   XL ALPHA AR.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93376782; PubMed=7690145;
RA   Fischer L., Catz D., Kelley D.;
RT   "An androgen receptor mRNA isoform associated with hormone-induced
RT   cell proliferation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:8254-8258(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95324753; PubMed=7601302;
RA   Fischer L.M., Catz D., Kelley D.B.;
RT   "Androgen directed development of the Xenopus laevis larynx: control
RT   of androgen receptor expression and tissue differentiation.";
RL   Dev. Biol. 170:115-126(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kelley D.B., Kamenetz F.K., Kelley D.B., Badea T.C.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; U67129; AAC97386.1; -.
DR   HSSP; P06536; 1GDC.
DR   InterPro; IPR001103; Andrgn_receptor.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF02166; Androgen_recep; 1.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   790 AA;  86973 MW;  9E7EC136CCCC3906 CRC64;

  Query Match             16.3%;  Score 456.5;  DB 13;  Length 790;
  Best Local Similarity   25.6%;  Pred. No. 2e-33;
  Matches  139;  Conservative   82;  Mismatches  170;  Indels  151;  Gaps   18;

Qy         18 CSQSILPLEHG-------------SIYIPSSYVDSHHEYPAMTFYS-------------- 50
              |:  ||  |              ::| |:::::   ||:  |||
Db        244 CTDLALPEPAGGYPHRAMELTPSLTLYKPTAFMEESPGYPSPDFYSPQMALAPHGRIKVF 303

Qy         51 ----------PAVMNYSIPSNVTN--------------------LEGGP----------G 70
                        |  ||  |   :                      | |,          |
Db        304 NPMFYGGGAWGAAGRYSFLSGFAHQGATAGWHTLFFFGQSSGSFAFAGPYSYPRSHGPAG 363

Qy         71 RQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEPQK---SPWCEA----------RSLEHT 116
                 |:  |    : :| :: |     |: | |   :|| |                :|
Db        364 ADGFFPSDAW----YPAPTMIGRV---PYSGPMKIEMAPWMEGYPGAFGEMPLFGGRDHL 416

Qy        117 IPVNPFTIKPKVSGNPCASPVTGPGSKPDAHP-----CAVCSDYASGYHYGVWSCEGCKA 171
              ||:                          | :|     | :| | ||| |||  :|  ||
Db        417 LPI-----------------------DYYFPPQKTCLICGDEASGCHYGALTCGSCKV 451

Qy        172 FFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRKERGGYRLVRR 231
              ||||: :|   |:| : | ||||| |||:| :|||||||| ||          | | :::
Db        452 FFKRAAEGKQKYLCASRNDCTIDKERRKNCPSCRLRKCYEAGMT--------IGAKKLKK 503

Qy        232 QRSADEQLHCAGKAKRSGCH   APEVRELLLDALSPEQLVLTLLEAFPPHVLIS--RPS 286
                :  |   | : :  |    :| :    |:  | : : | :|||  | |: :    :
```

```
Db        504 LGNLKAQEELDGSSVQGEGSKELSPGMGTPQLEGYSCQPIFLNVLEATEPVVVCAGHDNN 563

Qy        287 APFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCWMEVLMMGLMWR 346
              |  : |  :: || :| ::::|||:: ||| :||| | : ||: ::: || ::: : ||
Db        564 QPDSFALLLSSLNELGERQLVHVVKWAKALPGFRNLHVSDQMTVIQYSWMGLMIFAMGWR 623

Qy        347 SIDHPGK LIFAPDLVLD - - KDEGKCVEGILEIFDMLLATTSRFRELKLQHKEY 397
              |  :     | |||||| :       }   :||         |   :  |  |::  :|:
Db        624 SFKNVNSRMLYFAPDLVFNEYRMHKSRMYSQCVR--------LRHLSQEFGWLQITPEEF 675

Qy        398 LCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLAN 457
              ||:|||:|   |: | |   :|     :|         | ::     :  || ::  :|
Db        676 LCMKALLLF--STIP-VEGLKDQKCFDELRMNYIKELDRVISCKRNNPASSSRRFFQLTK 732

Qy        458 LL 459
              ||
Db        733 LL 734
```

Search completed: June 16, 2003, 14:15:13
Job time : 42.5411 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on June 15, 2003, 23:28:27 ; Search time 1913.05 Seconds
(without alignments)
10641.499 Million cell updates/sec

Title: US-08-826-361A-20
Perfect score: 1257
Sequence: 1 ATCAATTACAGCATTCCCAG..... GGCATGCGAGGTCTGCCTGA 1257

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: qb_est1:*
10: qb_est2:*
11: qb_htc:*
12: qb_est3:*
13: qb_est4:*
14: qb_est5:*
15: em_estfun:*
16: em_estom:*
17: qb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	262.4	20.9	727	12	BG722442	BG722442 602693671
2	162.2	12.9	580	12	BG738138	BG738138 fp04f05.y
3	157	12.5	642	12	BF058527	BF058527 7k31h08.x
4	155.4	12.4	641	10	AW976636	AW976636 ES1388745
5	153.8	12.2	648	12	BF660250	BF660250 maa95g12.
6	149.6	11.9	642	12	BF733032	BF733032 nae18e05.
c 7	143.4	11.4	627	14	BM942391	BM942391 UI-M-BH3-
8	143.2	11.4	782	10	AW134052	AW134052 fi14h07.y
9	131.4	10.5	594	12	BG084311	BG084311 H3099A08-
10	129.8	10.3	1142	14	BM810012	BM810012 AGENCOURT
11	126.4	10.1	137	17	AF455798	AF455798 AF455798
12	117	9.3	692	13	BJ068905	BJ068905 BJ068905
13	115.6	9.2	532	17	AZ596395	AZ596395 1M0409D10
14	115.2	9.2	611	17	AQ719425	AQ719425 HS_5512_B
15	110.6	8.8	1443	11	BC008218	BC008218 Homo sapi
16	108.8	8.7	610	13	BJ069238	BJ069238 BJ069238
17	107.8	8.6	703	12	BG744031	BG744031 602722811
c 18	107.4	8.5	575	12	BF510841	BF510841 UI-H-BI4-
19	107.4	8.5	2139	11	AK019223	AK019223 Mus muscu
20	106.2	8.4	676	12	BG747925	BG747925 602705778
21	102	8.1	844	13	BI100868	BI100868 602886184
22	101.4	8.1	937	13	BM562894	BM562894 AGENCOURT
23	100.4	8.0	558	13	BJ058831	BJ058831 BJ058831
24	99.8	7.9	908	10	BE546976	BE546976 601071950
25	99.6	7.9	800	9	AU169990	AU169990 AU169990
26	99.6	7.9	1004	9	AL527012	AL527012 AL527012
c 27	97.8	7.8	528	14	BM933826	BM933826 UI-M-BH3-
28	96.6	7.7	775	12	BG480960	BG480960 602529118
29	96.4	7.7	979	12	BG033384	BG033384 602301556
30	95	7.6	818	13	BI738331	BI738331 603361587
31	91.8	7.3	669	14	BM957249	BM957249 ty75b10.y
32	91.2	7.3	434	10	BB711891	BB711891 BB711891
33	91.2	7.3	521	12	BG687711	BG687711 334912 BA
34	91.2	7.3	698	10	BB196002	BB196002 BB196002
35	91.2	7.3	1638	11	BC016462	BC016462 Mus muscu
36	91	7.2	399	12	BG010872	BG010872 BC5-GN019
37	90.6	7.2	592	12	BG746447	BG746447 602703728
c 38	90.2	7.2	413	10	AW113877	AW113877 MC809 mou
39	90.2	7.2	1010	9	AL547737	AL547737 AL547737
40	89.6	7.1	275	10	BB196485	BB196485 BB196485
41	89.6	7.1	425	10	BB710863	BB710863 BB710863
42	89.2	7.1	1120	14	BM908889	BM908889 AGENCOURT
43	88.8	7.1	729	13	BG914551	BG914551 602813303
44	88.2	7.0	462	9	AA411285	AA411285 zv26g04.r
c 45	88.2	7.0	477	9	AI341113	AI341113 qx84c02.x
c 46	88.2	7.0	479	9	AI653325	AI653325 wb24c06.x
47	88.2	7.0	549	9	AL047238	AL047238 DKFZp586N
48	88.2	7.0	558	14	BM786552	BM786552 K-EST0065
c 49	88.2	7.0	644	12	BF223014	BF223014 7q27e02.x
50	88.2	7.0	754	12	BG682389	BG682389 602630127
51	88.2	7.0	889	9	AL561541	AL561541 AL561541
52	88.2	7.0	891	9	AL525589	AL525589 AL525589
53	88.2	7.0	909	14	BQ898227	BQ898227 AGENCOURT
54	88.2	7.0	926	9	AL513888	AL513888 AL513888
55	88.2	7.0	960	14	BQ879017	BQ879017 AGENCOURT
56	88.2	7.0	967	9	AL555839	AL555839 AL555839
57	88.2	7.0	972	14	BQ882678	BQ882678 AGENCOURT
58	88.2	7.0	1019	13	BM557938	BM557938 AGENCOURT
59	88.2	7.0	1090	12	BG176463	BG176463 602338147
60	86.8	6.9	674	14	BM791802	BM791802 K-EST0071
61	86.8	6.9	911	14	BQ948248	BQ948248 AGENCOURT
62	86.6	6.9	1176	12	BF162189	BF162189 601770663
63	86.4	6.9	385	10	AV647429	AV647429 AV647429
64	86.4	6.9	399	10	AV647483	AV647483 AV647483
65	86.4	6.9	399	10	AV647486	AV647486 AV647486
c 66	86.4	6.9	446	9	AI636264	AI636264 tz93e10.x
c 67	86.4	6.9	465	9	AI393892	AI393892 tq05e05.x
c 68	86.4	6.9	519	9	AU143800	AU143800 AU143800
69	86.4	6.9	616	12	BG723061	BG723061 602690412
70	86.2	6.9	863	14	BQ949290	BQ949290 AGENCOURT
71	86	6.8	416	13	BI133725	BI133725 UI-M-BH3-
72	86	6.8	794	12	BF138044	BF138044 601784742
c 73	85.8	6.8	271	14	BI300602	BI300602 UI-R-CV2-
74	85.8	6.8	450	17	AZ032574	AZ032574 RPCI-23-3
75	85.8	6.8	992	9	AL547687	AL547687 AL547687
76	85.8	6.8	1071	14	BM906677	BM906677 AGENCOURT
77	85.8	6.8	3141	11	BC030234	BC030234 Homo sapi
c 78	85.6	6.8	279	12	BG374797	BG374797 UI-R-CV1-
c 79	85.6	6.8	279	13	BI298646	BI298646 UI-R-CV2-

	80	85.6	6.8	569	10	AV968271
c	81	85.4	6.8	758	14	BQ446111
	82	85.2	6.8	645	10	BE533846
	83	85.2	6.8	947	9	AL527466
	84	85	6.8	910	9	AL556302
	85	85	6.8	2468	11	AK004709
c	86	84.8	6.7	345	12	BF551955
c	87	84.8	6.7	384	9	AA851319
c	88	84.8	6.7	384	9	AA893618
c	89	84.8	6.7	390	14	WI3715
	90	84.8	6.7	428	9	AI323943
c	91	84.8	6.7	451	9	AI327473
	92	84.8	6.7	500	10	AW490389
	93	84.8	6.7	574	9	AI171721
	94	84.8	6.7	588	10	AW491078
	95	84.8	6.7	655	10	AV647680
	96	84.8	6.7	963	14	BQ936063
	97	84.2	6.7	466	9	AA744644
c	98	84.2	6.7	610	14	BM941929
	99	84.2	6.7	621	12	BF143483
c	100	84.2	6.7	972	17	CNS04TUQ
c	101	84	6.7	276	13	BI300572
c	102	84	6.7	955	17	CNS04QE6
	103	83.8	6.7	783	9	AL561358
	104	83.8	6.7	1103	14	BM908460
	105	83.8	6.7	1114	12	BG031969
c	106	83.6	6.7	714	17	CNS01VNY
c	107	83.2	6.6	259	13	BI299473
	108	83.2	6.6	750	13	BI759107
	109	83.2	6.6	818	13	BM020032
	110	82.8	6.6	534	9	AA517920
	111	82.6	6.6	663	13	BI832037
	112	82.6	6.6	682	13	BI827058
	113	82.2	6.5	317	10	BB276737
	114	82.2	6.5	431	10	BE665969
	115	82.2	6.5	450	12	BF937961
	116	82.2	6.5	524	12	BG482331
	117	82.2	6.5	580	12	BG425812
	118	82.2	6.5	629	14	BM739741
	119	82.2	6.5	729	12	BF782180
	120	82.2	6.5	841	12	BG425174

AV968271	AV968271
BQ446111	UI-H-EU1-
BE533846	601233423
AL527466	AL527466
AL556302	AL556302
AK004709	Mus muscu
BF551955	UI-R-C2p-
AA851319	EST194087
AA893618	EST197421
WI3715	ma83b04.r1
AI323943	ma83b04.x
AI327473	ma83b04.y
AW490389	UI-M-BH3-
AI171721	EST217703
AW491078	UI-M-BH3-
AV647680	AV647680
BQ936063	AGENCOURT
AA744644	ny79f02.s
BM941929	UI-M-CG0p
BF143483	601790832
AL406971	Tetraodon
BI300572	UI-R-CV2-
AL402487	Tetraodon
AL561358	AL561358
BM908460	AGENCOURT
BG031969	602300663
AL169351	Tetraodon
BI299473	UI-R-CV2-
BI759107	603042512
BM020032	603648565
AA517920	vh95e10.r
BI832037	603075630
BI827058	603075729
BB276737	BB276737
BE665969	155233 MA
BF937961	tm70q06.y
BG482331	602526670
BG425812	602423886
BM739741	K-EST0010
BF782180	602106357
BG425174	602423729

ALIGNMENTS

```
RESULT 1
BG722442
LOCUS       BG722442                 727 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602693671F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4825937 5',
            mRNA sequence.
ACCESSION   BG722442
VERSION     BG722442.1  GI:14001629
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            %Plate: LLAM10739  row: n  column: 18
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1..727
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "IMAGE:4825937"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (qtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation). Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      182 a    194 c    175 g    176 t
ORIGIN

  Query Match             20.9%;  Score 262.4;  DB 12;  Length 727;
  Best Local Similarity   99.3%;  Pred. No. 1.7e-64;
  Matches  274;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 510

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 570

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 630

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAA-GCAAGTTAGTGGGAACCGTTC 239
              |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||
Db        631 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAACGGACGGTTAGTGGGAACCGTTG 690

Qy        240 CGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 275
              ||||||||||||||||||||||||||||||||||||
Db        691 CGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 726


RESULT 2
BG738138
LOCUS       BG738138                 580 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  fp04105.yl zebrafish gridded kidney Danio rerio cDNA clone 4728705
            5' similar to TR:O13012 O13012 ESTROGEN RECEPTOR. ;, mRNA sequence.
ACCESSION   BG738138
VERSION     BG738138.1  GI:14087827
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fp04105.xl
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
```

Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 477.

FEATURES Location/Qualifiers
source 1. .580
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="4728705"
/clone_lib "zebrafish gridded kidney"
/sex="mixed"
/tissue_type="kidney pooled from 300 wild type adults"
/lab_host="XLOLR"
/note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; Oligo dT cDNA library constructed from mRNA pooled from pooled kidney tissue from 300 adult zebrafish."

BASE COUNT 133 a 161 c 156 g 130 t
ORIGIN

Query Match 12.9%; Score 162.2; DB 12; Length 580;
Best Local Similarity 76.0%; Pred. No. 1.4e-35;
Matches 200; Conservative 0, Mismatches 63, Indels 0, Gaps 0,

```
Qy        268 AAGAGGGATGGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGA 327
              ||  | ||  |||||||||||||| || |||   || ||||| || || ||||||||||| 
Db        257 AAAGGAGACACTCACTTCTGCGCCGTGTGCCATGACTACGCTTCTGGGTATCACTATGGT 316

Qy        328 GTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGAT 387
              |||||||| |||||||| || ||||| || || ||  | ||||||||||| || ||||| 
Db        317 GTCTGGTCATGTGAAGGGTGCAAGGCTTTCTTCAAGCGTAGCATTCAAGCTCACAATGAC 376

Qy        388 TATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAG 447
              |||||||||||||| || || ||||| || || || || | | | |||||||||||| ||| 
Db        377 TATATTTGTCCAGCCACCAACCAGTGCACTATTGACAAGAGCAGACGCAAGAGCTGTCAG 436

Qy        448 GCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAG 507
              ||||| ||||| || ||||| || |||||||| ||| ||||||||||     |||| ||| 
Db        437 GCCTGTCGACTTCCGCAAGTGCTATGAAGTGGGCATGATGAAGTGTGGTGTTAGGAGGGAG 496

Qy        508 AGATGTCGGTACCGCCTTGTGCG 530
                | ||  | || ||   ||  ||
Db        497 CGCTGCAGTTATCGAGGTGCTCG 519
```

RESULT 3
BF058527
LOCUS BF058527 642 bp mRNA linear EST 16-OCT-2000
DEFINITION 7k31h08.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3477350 3' similar to SW:ESR1_HUMAN P03372 ESTROGEN RECEPTOR ;, mRNA sequence.
ACCESSION BF058527
VERSION BF058527.1 GI:10812423
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 492.

FEATURES Location/Qualifiers
source 1. .642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3477350"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note "Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGACATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 158 a 182 c 196 g 106 t
ORIGIN

Query Match 12.5%; Score 157; DB 12; Length 642;
Best Local Similarity 77.6%; Pred. No. 4.9e-34;
Matches 190; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||| 
Db        346 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 405

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| |||| || || |||||||| || || ||||| ||||||||||||||| || ||| 
Db        406 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 465

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || ||||||||| || | ||||||||||||||| 
Db        466 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 525

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  || 
Db        526 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGCTGGGATACGAAAAGACCGA 585

Qy        511 TGTGG 515
                | ||
Db        586 AGAGG 590
```

RESULT 4
AW976636
LOCUS AW976636 641 bp mRNA linear EST 02-JUN-2000
DEFINITION EST388745 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION AW976636
VERSION AW976636.1 GI:8167867
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 641)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research

```
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 363
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1..641
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGN"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      159 a    175 c    195 g    112 t
ORIGIN

  Query Match             12.4%;  Score 155.4;  DB 10;  Length 641;
  Best Local Similarity   77.1%;  Pred. No. 1.4e-33;
  Matches  189;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        334 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 393

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        394 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 453

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || ||||||||| || | ||||||||||||||||
Db        454 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 513

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || |||||||||||||||| |||| ||||    || | |||  ||
Db        514 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 573

Qy        511 TGTGG 515
                ||
Db        574 ATAGG 578


RESULT 5
BF660250
LOCUS       BF660250                 648 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  maa95q12.x1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3824830 3' similar to SW:ESR1_MOUSE P19785 ESTROGEN RECEPTOR
            ;, mRNA sequence.
ACCESSION   BF660250
VERSION     BF660250.1  GI:11925456
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1461182
            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1..648
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3824830"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      158 a    184 c    183 g    123 t
ORIGIN

  Query Match             12.2%;  Score 153.8;  DB 12;  Length 648;
  Best Local Similarity   76.7%;  Pred. No. 4.1e-33;
  Matches  188;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || ||||| |||
Db        283 AAGGAGACTCGCTACTGTGCCGTGTGCAATGACTATGCCTCTGGCTACCATTATGGGGTC 342

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| || ||||| || ||||| || ||||| ||||||||||||||| ||||| ||
Db        343 TGGTCCTGCGAAGGCTGCAAGGCTTTCTTTAAGAGAAGCATTCAAGGACACAATGACTAC 402

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||||||||| || || || || || || |||||| | ||||| |||||||||
Db        403 ATGTGTCCAGCTACAAACCAATGCACCATTGACAAGAACCGGAGGAAGAGTTGCCAGGCC 462

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              || || || || ||||||||||||||||| ||| |||| | |||  |||| ||| |
Db        463 TGTCGGCTGCGCAAGTGTTACGAAGTGGGCATGATGAAAGGCGGCATACGGAAAGACCGC 522

Qy        511 TGTGG 515
                | ||
Db        523 CGAGG 527


RESULT 6
BF733032
LOCUS       BF733032                 642 bp    mRNA    linear   EST 30-MAR-2001
DEFINITION  nae18e05.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3435729 3'
            similar to SW:ESR1_HUMAN P03372 ESTROGEN RECEPTOR ;, mRNA sequence.
ACCESSION   BF733032
VERSION     BF733032.1  GI:12058107
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1..642
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3435729"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note "Organ: ovary, Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker, Site_1: Not I, Site_2: Eco RI, 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGACAIIIIIIIIIIIIIIIIIII 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 156 a 184 c 193 g 109 t
ORIGIN

Query Match 11.9%; Score 149.6; DB 12; Length 642;
Best Local Similarity 77.1%; Pred. No. 6.7e-32;
Matches 182; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 271 AGGCATCCTCACTTCTCGCCTCTCTCCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
Db 346 AAGGAGACTCGCTACCTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 405

Qy 331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
Db 406 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 465

Qy 391 ATTTGTCCAGCTACAAATCAGTGTACAATGGATAAAAACTGGCGCAAGAGCTGCCAGGCC 450
Db 466 ATGTGTCCAGCCACCAACCAGTGCAACATTGATAAAAACAGGAAGGAAGACTGTCAAGCC 525

Qy 451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGA 506
Db 526 TGCCGGCTCGGCAAATGCTACGAAATGGGAATGATGAAAGGTGGGATACGATAAGA 581

RESULT 7
BM942391/c
LOCUS BM942391 627 bp mRNA linear EST 12-MAR-2002
DEFINITION UI-M-BH3-bsc-b-09-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone UI-M-BH3-bsc-b-09-0-UI 5', mRNA sequence.
ACCESSION BM942391
VERSION BM942391.1 GI:19402254
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 627)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 REVERSE.
FEATURES Location/Qualifiers
source 1. .627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-bsc-b-09-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1 The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)"

BASE COUNT 147 a 151 c 184 g 144 t 1 others
ORIGIN

Query Match 11.4%; Score 143.4; DB 14; Length 627;
Best Local Similarity 59.7%; Pred. No. 4.1e-30;
Matches 259; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

Qy 817 CACCAAGTCCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGG 876
Db 624 GATCAGGTCCACCTTCTCGACTGTGCCTGGCTGCAGATTCTGATGATTGGTCTCGTCTGG 565

Qy 877 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGAT 936
Db 564 CGCTCCATGGAACACCCCGAGAAGCTCCTGTNTGCTCCTAACTTGCTCCTGGACAGGAAT 505

Qy 937 GAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCA 996
Db 504 CAAGCTAAATGTGTGGAAGGCATGGTGGAGATCTTTGACATCTTGCTGGCTACGTCAAGT 445

Qy 997 AGCTTTCCAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1056
Db 444 CGGTTCGGCATGATGAACCTGCAGGGAGAAGAGTTTGTGTGCCTCAAATCCATCATTTTG 385

Qy 1057 CTCAATTCCACTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGG 1113
Db 384 CTTAATTCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGAC 325

Qy 1114 AAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGC 1173
Db 324 CACATCCACCGTGTCCTGGACAAGATCACAGACACTTTGATCCAGCTGATGGCCAAAGCT 265

Qy 1174 GGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCAC 1233
Db 264 GGCCTGACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCAT 205

Qy 1234 GTCAGCCATCCGAG 1247
Db 204 ATCCGGCACATGAG 191

RESULT 8

```
AW134052
LOCUS       AW134052                782 bp    mRNA    linear   EST 27-OCT-1999
DEFINITION  fi14h07.y1 Suqano Kawakami zebrafish DRA Danio rerio cDNA clone
            2601181 5' similar to qb:M12674 ESTROGEN RECEPTOR (HUMAN);, mRNA
            sequence.
ACCESSION   AW134052
VERSION     AW134052.1  GI:6135659
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopteryqii; Neopteryqii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 782)
  AUTHORS   Suqano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theisinq,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     other_ESTs: fi14h07.x1
            Contact: S.L. Johnson
            Washinqton University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Suqano and Dr. Koichi Kawakami DNA
            Sequencinq by: Washinqton University Genome Sequencinq Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1. .782
                     /orqanism "Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="2601181"
                     /clone_lib="Suqano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, includinq
                     unfertilized eqqs)"
                     /dev_staqe "adult"
                     /lab_host "DH10B (phaqe resistant)"
                     /note "Vector: pMEI8S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oliqo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was liqated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], diqested and cloned into distinct DraIII
                     sites of the pMEI8S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fraqments
                     <1.5kb.  Library constructed by Dr. Sumio Suqano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami.  Custom primers for
                     sequencinq: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA.  "
BASE COUNT      195 a    230 c    176 q    179 t      2 others
ORIGIN
```

```
  Query Match             11.4%;  Score 143.2;  DB 10;  Length 782;
  Best Local Similarity   61.0%;  Pred. No. 5.8e-30;
  Matches  267;  Conservative    0;  Mismatches  165;  Indels    6;  Gaps    2;

Qy          18 CAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTT 77
               ||||||      ||   | |||  |   | |||  || ||| |   |||||| | | | ||
Db         351 CAGCACATCGCCTTTATCGGACTGCTCGTCTGTGCGCCAGTCGCTAAGCCCGACTTTATT 410

Qy          78 GTGGCCAACACCTGGGCACCTTTCTCTCTTTAGTCGTCCATCGCCAGTTATCACATCTGTA 137
               |||||| | |  | ||  |||| | |     | || | |||     || ||   |
Db         411 TTGGCCACCTCACAGCCATGTTTCCTCACTCACATTGCAACAACAGAGTCGACTTCAACA 470

Qy         138 TGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGT 197
                 | ||   |  | | |||   |||   | | | | || ||     |   |
Db         471 GAACCATGCTACTAGTGGGACTTGGACAGAACATACACCACATGATCATGTTGAAGAGGA 530

Qy         198 AAACAGAGAGACACTGAAAAGCAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGC 257
               |||||| | |||||  | | || ||| |  || | | | |  | |||  |||
Db         531 AAACAGCAAACCACTGGTGAAGCGGG-TAGCAGATACAGAGGAGACGTCTGTGTCTCTGA 589

Qy         258 TCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATA 317
                   |   ||    ||    |||| ||| || |||||||| |||||||| || |||||
Db         590 -----GAGGTAAAGCTGACATGCACTACTGTGCCGTCTGCAGTGATTACGCCTCTGGATA 644

Qy         318 TCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGG 377
               ||| |||||||| |||||||||||||| || || |||||||| || || || || |||||
Db         645 TCATTATGGAGTGTGGTCGTGTGAAGGCTGCAAAGCCTTTTTCAAGAGGAGTATACAAGG 704

Qy         378 ACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAA 437
               ||| ||||| || || ||||| || ||||| || || || || |||||  | ||||||||||
Db         705 ACACAATGACTACATCTGTCCTGCCACAAACCAATGCACTATCGANCAGAACCGGCGCAN 764

Qy         438 GAGCTGCCAGGCCTGCCG 455
               ||||||||||||||||||
Db         765 AAGCTGCCAGGCCTGCCG 782
```

```
RESULT 9
BG084311
LOCUS       BG084311                594 bp    mRNA    linear   EST 26-JAN-2001
DEFINITION  H3099A08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3099A08 5', mRNA sequence.
ACCESSION   BG084311
VERSION     BG084311.1  GI:12566875
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Karqul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3099A08-3
            Contact: George J. Karqul
            Laboratory of Genetics
            National Institute on Aqinq/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lqsun.qrc.nia.nih.qov
            This clone set has been freely distributed to the community. Please
            visit http://lqsun.qrc.nia.nih.qov/cDNA/15k.html for details.
            Plate: H3099  row: A  column: 08
            Seq primer: -21M13 Reverse
            High quality sequence stop: 594
            POLYA No.
FEATURES             Location/Qualifiers
     source          1. .594
                     /orqanism="Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref "nia EST:H3099A08-5"
                     /db_xref="taxon:10090"
                     /clone="H3099A08"
                     /clone_lib "NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_staqe="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host "DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is amonq a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (includinq preimplantation staqe
                     embryos from  unfertilized eqq to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/qonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with oliqo(dT
                     )-Not primers.  References include: (1) Genome-wide
                     expression profilinq of mid-qestation  placenta and embryo
```

using a 15,000 mouse developmental cDNA microarray. 2000. Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
BASE COUNT 170 a 124 c 167 g 133 t
ORIGIN

Query Match 10.5%; Score 131.4; DB 12; Length 594;
Best Local Similarity 79.2%; Pred. No. 1.2e-26;
Matches 156; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy        319 CACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGA 378
              || ||||| |||||||| || ||||| || ||||| || ||||| ||||||||||||||||
Db          1 CATTATGGGGTCTGGTCCTGCGAAGGCTGCAAGGCTTTCTTTAAGAGAAGCATTCAAGGA 60

Qy        379 CATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAG 438
              || ||||| || || |||||||||||||| || || || || || || || |||||| | |||
Db         61 CACAATGACTACATGTGTCCAGCTACAAACCAATGCACCATTGACAAGAACCGGAGGAAG 120

Qy        439 AGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCC 498
              || |||||||||| || || || |||||||||||||||| ||| |||| | |||
Db        121 AGTTGCCAGGCCTGTCGGCTGCGCAAGTGTTACGAAGTGGGCATGATGAAAGGCGGCATA 180

Qy        499 CGGAGAGAGAGATGTGG 515
              |||| ||| | | ||
Db        181 CGGAAAGACCGCCGAGG 197
```

RESULT 10
BM810012
LOCUS BM810012 1142 bp mRNA linear EST 25-MAR-2002
DEFINITION AGENCOURT_6581896 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454957 5', mRNA sequence.
ACCESSION BM810012
VERSION BM810012.1 GI:19126835
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1142)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1950 row: o column: 22
High quality sequence start: 13
High quality sequence stop: 694.
FEATURES Location/Qualifiers
source 1..1142
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5454957"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 191 a 334 c 398 g 215 t 4 others
ORIGIN

Query Match 10.4%; Score 129.8; DB 14; Length 1142;
Best Local Similarity 48.7%; Pred. No. 5e-26;
Matches 381; Conservative 0; Mismatches 399; Indels 3; Gaps 1;

```
Qy        282 CTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGA 341
              | ||||| |||||  | ||    || || || || ||||||| ||    || ||||
Db         44 CCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGTGA 103

Qy        342 ACGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGC 401
              | || || ||||| || || || | ||| || ||    |  || || |  ||||| ||
Db        104 GGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCGGC 163

Qy        402 TACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCG 461
              | ||  ||||  |||  ||    | ||||||  ||||||||||||||  |
Db        164 CTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCTGCCAGGCCTGCCGCTTCAC 223

Qy        462 GAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCGGCAGAGAGATGTGGCTACCC 521
              |||||       ||||| ||| | |||   ||    ||    ||  |    ||   |
Db        224 CAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGGCTGGACCGCGTCCGGGGTGG 283

Qy        522 CCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAG 581
              |   |  |    || | |  | | || | | | |  |    |  ||
Db        284 GCGGCAGAAGTACAAGCGGCGGCGGAGGTGGACCCACTGCCCTTCCCGGGCCCTTCCC 343

Qy        582 AAGTGGCGGCCACGGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCA 641
              |||  ||  ||   |  |  |  | | |  |  |||||       |  |
Db        344 TGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACT 403

Qy        642 GCTACTGCTCACCTCCTGCAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGC 701
              | |    |   ||  ||| |    |  |        ||      ||||   |
Db        404 GGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGG 463

Qy        702 GCCCTTCAGCGAGGCCTGCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGT 761
              ||        |  | |  ||   ||||    | |   |||   ||| | ||
Db        464 CCCTGATGGGCACCTGCCAGCCGTGGCTACCCTCTGTGACCCTCTTTGACCGAGAGATTGT 523

Qy        762 ACACATCATCACGTGCCCCAAGAACATTCCCGGCTTTGTGGACCTCAGCCTGTTCACCA 821
              ||  |||||||||||||||  || || |||||      |||    ||||  |||||
Db        524 GGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCA 583

Qy        822 AGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTC 881
              ||    |  || |||||  |||||||||||| |  || ||||  ||      ||||||
Db        584 GATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGCTC 643

Qy        882 AATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGG 941
              | |    |  | |   ||||  ||| |||   ||  | || ||||    || ||||
Db        644 ACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTAGTCCTGGA---TGAAGAGGG 700

Qy        942 GAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCTTGGCAACTACTTCAAGGTT 1001
              | | | | ||  ||  |  |||| |    |     || |||  |           || |
Db        701 GGCAGGGCAAGTGGCCTGNGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGGCGCT 760

Qy       1002 TCGAGACTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAA 1061
              | |   |    ||  | |  |||| ||| |     |  |||| ||  |||   |||
Db        761 GCAGGCCCTGCGGCTGGAGCGAGAAGAGTATGTTCTACTANAGGCCTTGGCCCTTGCCAA 820

Qy       1062 TTC 1064
              |||
Db        821 TTC 823
```

RESULT 11
AF455798
LOCUS AF455798 137 bp DNA linear GSS 23-APR-2002
DEFINITION AF455798 Macaca mulatta DNA (Murphy WJ) Macaca mulatta genomic clone ESR2, DNA sequence.

```
ACCESSION   AF455798
VERSION     AF455798.1  GI:20271896
KEYWORDS    GSS.
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 137)
  AUTHORS   Murphy,W.J., Page,J.E., Smith,C. Jr., Desrosiers,R.C. and O'Brien
            ,S.J.
  TITLE     A radiation hybrid mapping panel for the rhesus macaque
  JOURNAL   J. Hered. 92 (6), 516-519 (2001)
  MEDLINE   21945527
COMMENT     Contact: Murphy WJ
            Laboratory of Genomic Diversity
            National Cancer Institute FCRDC
            P.O. Box B, Bldg. 560, Rm.11-10, Frederick, MD 21702, USA
            Tel: 301 846 1299
            Fax: 301 846 6327
            Email: murphywj@mail.ncifcrf.gov
            Class: unknown
FEATURES             Location/Qualifiers
     source          1..137
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /clone="ESR2"
                     /clone_lib="Macaca mulatta DNA (Murphy WJ)"
BASE COUNT       34 a      40 c      30 g      33 t
ORIGIN

  Query Match             10.1%;  Score 126.4;  DB 17;  Length 137;
  Best Local Similarity   95.6%;  Pred. No. 1.4e-25;
  Matches  130;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          5 ATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAA 64
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          1 ATTACAGCAGTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCGCAA 60

Qy         65 GCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGT 124
              ||||||||||||||| |||||||||||| |||||||||||||||  |||||||||||||
Db         61 GCCCAAATGTGTTGTGGCCAACACCTGGACACCTTTCTCCTTTAGCAGTCCATCGCCAGT 120

Qy        125 TATCACATCTGTATGC 140
              |||||||||| |||||
Db        121 TATCACATCTATATGC 136


RESULT 12
BJ068905
LOCUS       BJ068905                692 bp    mRNA    linear   EST 11-DEC 2001
DEFINITION  BJ068905 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL051b20 5', mRNA sequence.
ACCESSION   BJ068905
VERSION     BJ068905.1  GI:17497655
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559 81-6856
            Fax: 81 559 81 6855
            Email: tshini@genes.nig.ac.jp.

FEATURES             Location/Qualifiers
     source          1..692
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL051b20"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
                     /note="Vector: pBSKN+; Site_1: NotI; Site_2: EcoKI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett and A.M. Zorn,
                     (Wellcome/CRC Institute). "
BASE COUNT      160 a     168 c     182 g     182 t
ORIGIN

  Query Match             9.3%;  Score 117;  DB 14;  Length 692;
  Best Local Similarity   51.6%;  Pred. No. 1.9e-22;
  Matches  329;  Conservative    0;  Mismatches  290;  Indels   18;  Gaps    2;

Qy        305 ACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAA 364
              | ||||| || ||||||||||| ||    || ||||| | ||| || ||||| || ||
Db         10 AGGCATCAGGCTATCACTATGGGGTTGCCTCATGTGAGGCATGCAAAGCCTTCTTCAAGC 69

Qy        365 GAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATA 424
              | | ||| ||||| ||| || || | |||||||| | ||| | |||  |||    |
Db         70 GCACCATCCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATGAATGTGAAATTACCA 129

Qy        425 AAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGG 484
              | |  || || ||  ||| || || ||||  |   |||||       ||||| |||
Db        130 AGAGACGTCGAAAGGCCTGTCAAGCTTGCCGCTTCACAAAGTGCCTTCGCGTGGGCATGC 189

Qy        485 TGAAGTGTGGCTCCCGGAGAGACAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTG 544
              | ||   ||    ||    || |     |     | | | |       |||| ||
Db        190 TAAAAGAAGGGGTACGCTTGGACCGTGTCAGAGGAGGACGTCAGAAATACAAGAGGAGGC 249

Qy        545 CCGACGACGACGTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCCCGAG 604
              | || |   |  | | |                   ||| || || || |  | ||
Db        250 CAGATGGAGATATCCTC- ------- --CAGTACTCGTCTGGTGGGCAGGCTCAGCAAG 297

Qy        605 TGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGG 664
                  |     || |  |    |  |  || | | |   |||  || | | ||||  ||
Db        298 CCTCTGCAACAGTGGTCAAGAAGCAAACTCCAGTGAATGCAGTTGTCTCTCACCTGCTGG 357

Qy        665 CTGAGCCGCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATG 724
               ||     ||  |   | || |       ||     |  |  |  | || |       ||
Db        358 TTGCAGAACCAGACAAGTTGTTTGCTATGCCAGACCCAGCACTTCCTGATGGTGATTTGA 417

Qy        725 -----ATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGG 778
                  ||||  |||||    |  | || ||    ||  | ||   ||| || |||||||
Db        418 AATATATGAGCACCCTGTGTGACCTTGCTGATCGTGAAATTGTCATCATCATTAGCTGGG 477

Qy        779 CCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGA 838
              | |||| || || || ||          ||   ||| |  |||||  || | |||||| |||
Db        478 CAAAGAATATCCCAGGATTCTCCTCCCTTTCCCTCTCGGACCAGATGAGTCTCTTGCAGA 537

Qy        839 GCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCA 898
              |    |||||||||||| ||| || | ||  |  | |  ||||  ||    |    |
Db        538 GTGTCTGGATGGAGGTTTTATTGCTTGGAGTTGTTTCCGCTCGCTTCCGTATGAAGATG 597

Qy        899 AGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGA 935
              || |  | ||||||   ||| |||| || ||   |||
Db        598 AGGTTGTGTTTGCTGAGGATTTTGTGCTTGATGAGGA 634


RESULT 14
AZ596395
LOCUS       AZ596395                532 bp    DNA     linear   GSS 13-DEC 2000
DEFINITION  1M0409D10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0409D10 R, DNA sequence.
ACCESSION   AZ596395
VERSION     AZ596395.1  GI:11718585
```

```
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Dunn,D., Aoyaqi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Lonqacre,S., Mahmoud,M., Meenen,E., Pedersen,I., Reilly
            ,M., Rose,M., Rose,K., Stokes,R., Tinqey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole qenome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldq., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@qenetics.utah.edu
            Insert Lenqth: 10000  Std Error: 0.00
            Plate: 0409  row: D  column: 10
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 532.
FEATURES             Location/Qualifiers
     source          1. .532
                     /orqanism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M04D9D10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F'"
                     /note="Vector: PWD42nv; Purified qenomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/Shares/).  The DNA
                     was hydrodynamically sheared by repeated passaqe through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oliqonucleotides were
                     liqated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb rande using preparative agarose qel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (qi|4732114|qb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was liqated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically competent E. coli XL10 Gold (Stratadene) cells
                     and selected for ampicillin resistance."
BASE COUNT       118 a    132 c    135 q    147 t
ORIGIN

  Query Match             9.2%;  Score 115.6;  DB 17;  Lenqth 532;
  Best Local Similarity   84.4%;  Pred. No. 4.1e-22;
  Matches  130;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        930 CAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAAC 989
              ||||||||||||||| ||||| ||||| ||||||||||||||||||||||||||||||||
Db        257 CAGGGATGAGGGGAAGTCCGTGGAAGGCATTCTGGAAATCTTTGACATGCTCCTGGCGAC 316

Qy        990 TACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCAT 1049
              ||  || |||| || |||||||||| || |||||||||||| ||||| |||||||
Db        317 GACGGCACGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCAT 376

Qy       1050 GATCCTGCTCAATTCCAGTATGTACCCTCTGGTC 1083
              ||| || ||||| |||||| ||| || | | |
Db        377 GATTCTGCTCAACTCCAGTGAGTAGCCACGAGGC 410
```

```
RESULT 14
AQ719425
LOCUS       AQ719425                611 bp    DNA     linear   GSS 14-JUL-1999
DEFINITION  HS_5512_B2_D07_T7A RPCI-11 Human Male BAC Library Homo sapiens
            qenomic clone Plate=1088 Col=14 Row=H, DNA sequence.
ACCESSION   AQ719425
VERSION     AQ719425.1  GI:5479094
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagqed connectors: A sequence approach to mapping and
            scanninq the human qenome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencinq Center
            University of Washinqton
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washinqton.edu
            Clones are derived from the human BAC library RPCI 11. For BAC
            library availability, please contact Pieter de Jonq
            (pieter@dejonq.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resqen.com). BAC end Web Server:
            http://www.htsc.washinqton.edu
            Plate: 1088  row: H  column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 611.
FEATURES             Location/Qualifiers
     source          1. .611
                     /orqanism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate:1088 Col:14 Row:H"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially diqested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       178 a    141 c    129 q    144 t     19 others
ORIGIN

  Query Match             9.2%;  Score 115.2;  DB 17;  Lenqth 611;
  Best Local Similarity   97.5%;  Pred. No. 5.8e-22;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        375 AGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCG 434
              | |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        337 AAGACATAATGATGATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCG 396

Qy        435 CAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 494
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        397 CAAGAGCTGCCAGGCCTGCCGACTTCTGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 456


RESULT 15
BC008218
LOCUS       BC008218                1443 bp    mRNA    linear   HTC 12-JUL-2001
DEFINITION  Homo sapiens, clone IMAGE:3871297, mRNA.
ACCESSION   BC008218
VERSION     BC008218.1  GI:14713363
KEYWORDS    HTC.
```

```
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1443)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 13 Row: k Column: 10
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4335517
            This clone has the following problem: incomplete processing.
FEATURES             Location/Qualifiers
     source          1..1443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4871297"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_67"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6."
BASE COUNT       490 a    339 c    352 g    362 t
ORIGIN

  Query Match             8.8%;  Score 110.6;  DB 11;  Length 1443;
  Best Local Similarity   49.4%;  Pred. No. 2e-20;
  Matches  342;  Conservative    0;  Mismatches  424;  Indels   27;  Gaps    1;

Qy        292 CTCTGCAGCGATTACGCATCGGGATATCACTATCGAGTCTGGTCGTGTGAAGGATGTAAG 351
              || ||  | ||   ||| || || || ||||||||| ||    || |||||||  || |||
Db        732 GTGTGTGGTGACATCGTTCTGGGTACCACTATGGGTAGCATCATGTGAAGCCTGCAAG 791

Qy        352 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 411
              || || || || || |  |||||||  |||  || || |  || || || || ||| |
Db        792 GCATTCTTCAAGAGGACAATTCAAGGCAATATAGAATACAGCTGCCCTGCCACGAATGAA 851

Qy        412 TCTACAATCGATAAAAACCGGCGCCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTAC 471
              |||  ||||   ||   | | || ||   |||||||| ||||  |   ||||||||
Db        852 TGTGAAATCACAAAGCGCAGACGTAAATCCTGCCAGGCTTGCCGCTTCATGAAGTGTTTA 911

Qy        472 GAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGC 531
               |||||||  ||| ||||    ||    ||    || |||
Db        912 AAAGTGGGCATGCTGAAAGAAGGGGTGCGTCTTGACAGA--       ---              950

Qy        532 AGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGC 591
                  | ||| |   |||| | ||     | | | |   |   |   |   ||
Db        951 --  --GTACGTGGAGGTCGGCAGAAGTACAAGCGCAGGATAGATGCCGGAGAACAGCCCA 1004

Qy        592 CACCCGCCCCGAGTGCCCCAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTC 651
              ||  |  ||    || ||  | ||       |       |  |||       |    |
Db       1005 TACCTGAACCCTCAGCTGGTTCAGCCAGCCAAAAAGCCATATAACAAGATTGTCTCACAT 1064

Qy        652 ACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACC 711
```

```
              |  |||  |    |  |             ||      ||| | ||  |||  ||
Db       1065 TTGTTGGTGGCTGAACCGGAGAAGATCTATGCCATGCCTGACCCTACTGTCCCCGACAGT 1124

Qy        712 GAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATC 771
              ||   |       | |   | |||    | |||||||||   ||||||||    || ||
Db       1125 GACATCAAAGCCCTCACTACACTGTGTGACTTGGCCGACCGAGAGTTGGTGGTTATCATT 1184

Qy        772 AGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTC 831
              | ||||| ||| | ||||| |||||       |||   ||||   |||||  || | ||
Db       1185 GGATGGGCGAAGCATATTCCAGGCTTCTCCACGCTGTCCCTGGCGGACCAGATGAGCTT 1244

Qy        832 TTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCAC 891
              || ||||   |||||||||  | || ||  | ||  |  | |  || ||  ||
Db       1245 CTGCAGAGTGCTTGGATGGAAATTTTGATCCTTGGTGTCGTATACCGGTCTCTTTCGTTT 1304

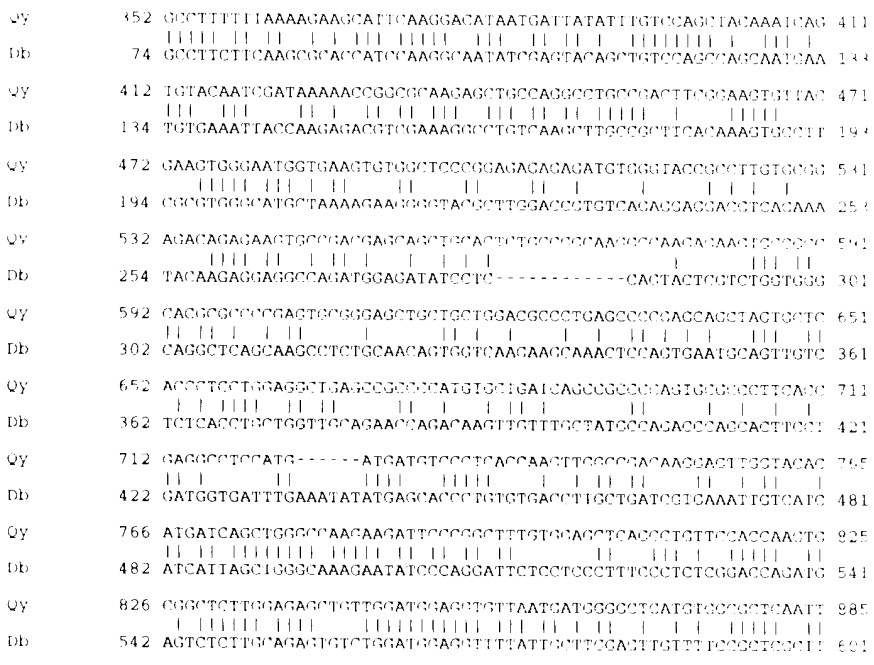
Qy        892 CCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTA 951
              |   | ||  ||| |||    ||| | |  ||||   || ||  ||  |||| | |
Db       1305 GAGGATGAACTTGTCTATGCAGACGATATATAATGCACGAAGACCAGTCCAAATTAGCA 1364

Qy        952 GAAGGAATTCTGGAAATCTTTGACATGCTCCTG 984
              |      ||     ||    ||  || || | |
Db       1365 GGCCTTCTTGATCTAAATAATGCTATCCTGCAG 1397


RESULT 16
BJ069238
LOCUS       BJ069238                 610 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ069238 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL052d06 5', mRNA sequence.
ACCESSION   BJ069238
VERSION     BJ069238.1  GI:17494935
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..610
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL052d06"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
                     /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett  and A.M. Zorn,
                     (Wellcome/CRC Institute). "
BASE COUNT       149 a    144 c    165 g    151 t      1 others
ORIGIN

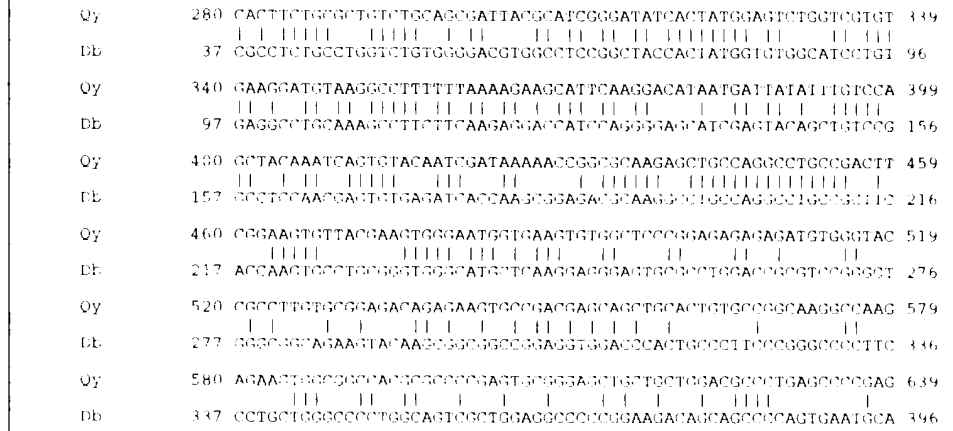
  Query Match             8.7%;  Score 108.8;  DB 13;  Length 610;
  Best Local Similarity   51.7%;  Pred. No. 4.1e-20;
  Matches  318;  Conservative    0;  Mismatches  272;  Indels   18;  Gaps    2;

Qy        292 CTCTGCAGCGATTACGCATCGGGATATCACTATGCAGTCTGGTCGTGTGAAGGATGTAAG 351
              || ||  | |||   ||||| || |||||||||||| ||    || |||| | ||| ||
Db         14 GTGTGTGGAGATGTGGCATCAGGCTATCACTATGGGGTTGCCTCATGTGAGGCATGCAAA 73
```



RESULT 17
BG744031
LOCUS BG744031 703 bp mRNA linear EST 15-MAY-2001
DEFINITION 6027228111F1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849155 5',
mRNA sequence.
ACCESSION BG744031
VERSION BG744031.1 GI:14054684
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 703)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1688 row: f column: 04
High quality sequence stop: 628.
FEATURES Location/Qualifiers
source 1..703
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4849155"
/clone_lib="NIH_MGC_106"
/tissue_type="natural killer cells, cell line"
/lab_host "DH10B (phage-resistant)"
/note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 116 a 210 c 245 g 132 t
ORIGIN

Query Match 8.6%; Score 107.8; DB 12; Length 703;
Best Local Similarity 48.6%; Pred. No. 8.7e-20;
Matches 295; Conservative 0; Mismatches 312; Indels 0; Gaps 0;



RESULT 18
BF510841/c
LOCUS BF510841 575 bp mRNA linear EST 06-DEC-2000
DEFINITION UI-H-BI4-apo-e-04-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3088206 3', mRNA sequence.
ACCESSION BF510841
VERSION BF510841.1 GI:11594139
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)

```
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaphs r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M B  Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html The following repetitive
            elements were found in this cDNA sequence: 1-48,
            -AT_rich#Low_complexity 449-568, -LINE2
            Seq primer: M13 Forward
            POLYA Yes.
FEATURES             Location/Qualifiers
     source          1..575
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:3088206"
                     /clone_lib "NCI_CGAP_Sub8"
                     /lab_host "DH10B (Life Technologies)"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
                     is a subtracted library derived from NCI_CGAP_Sub5. The
                     NCI_CGAP_Sub8 library had 2.5 million recombinants. A
                     single-stranded DNA preparation of NCI_CGAP_Sub5 was used
                     as a tracer in a subtractive  hybridization with a driver
                     comprising: a pool of clones from  NCI_CGAP_Sub5 (IMAGE
                     clone Ids: 2732843-2737415, 3068040-3069191, 25% of the
                     driver population), a pool of clones from NCI_CGAP_Sub4
                     (IMAGE clone Ids 2728592-2729426; 25% of the driver
                     population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids
                     2728969-2734190; 25% of the driver population), and
                     NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238,  3081864-3084550
                     ; 25% of the driver population).  Subtraction was
                     performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To Facilitate Gene Discovery. Genome Research
                     6, 791-806.
                     TAG_LIB NCI_CGAP_GC4
                     TAG_TISSUE germ cell
                     TAG_SEQ AAATC"
BASE COUNT       126 a    141 c    115 g    203 t
ORIGIN

  Query Match             8.5%;  Score 107.4;  DB 12;  Length 575;
  Best Local Similarity   70.2%;  Pred. No. 1e-19;
  Matches  144;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         10 AGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCA 69
              || ||||   |  ||| |   |    | |||   | ||    | |||   ||
Db        447 AGAATTCAAGGAGATGAGAAGGAAACAGGAGGGAAGATTGAAAAGAAGAGTCCAGTGTGT 388

Qy         70 AATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCA 129
                ||| |      || |       ||  | ||       ||||||||||||||||||||
Db        387 TATGAGGAAAACCCCAAGAGCATGCTGCCTTACAAGACAGTGGTCCATCGCCAGTTATCA 328

Qy        130 CATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACC 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 CATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACC 268

Qy        190 TTACCTGTAAACAGAGAGACACTGA 214
              |||||||||||||||||||||||||
Db        267 TTACCTGTAAACAGAGAGACACTGA 243


RESULT 19
AK019222
LOCUS       AK019222                2199 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus 11 days embryo whole body cDNA, RIKEN full length
            enriched library, clone:2700080M08 estrogen related receptor,
            alpha, full insert sequence.
ACCESSION   AK019222
VERSION     AK019222.1  GI:12859415
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2700080M08.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap trapper selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384 format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5  (bases 1 to 2199)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,F., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
```

```
            Sogabe,Y., Suzuki,H., Taqami,M., Taqawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81 45 503 9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..2199
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2700080M08"
                     /db_xref="MGD:MGI:1899528"
                     /db_xref="taxon:10090"
                     /clone="2700080M08"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="11 days embryo"
     gene            1..2199
                     /gene="Esrra"
     misc_feature    1..2199
                     /gene="Esrra"
                     /note="data source:MGD; source key:MGI:1346831
                     evidence:ISS
                     estrogen related receptor, alpha"
                     /db_xref="MGD:MGI:1346831"
BASE COUNT      425 a    622 c    722 g    430 t
ORIGIN

  Query Match             8.5%;  Score 107.4;  DB 11;  Length 2199;
  Best Local Similarity   48.7%;  Pred. No. 2.2e-19;
  Matches  322;  Conservative    0;  Mismatches  336;  Indels    3;  Gaps    1;

Qy        280 CACTTCTGCGCTGTCTGCAGCGATTACCCATCCGGATATCACTATGGAGTCTGGTCCTCTCT 339
              | | |||||   |||||  | |||   || || || || ||||| || ||    || |||
Db        389 CGCCTCTGCCTGGTCTGTGGGGATGTGGCCTCTGGCTACCACTACGGTGTGGCATCCTGT 448

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db        449 GAGGCCTGCAAACCCTTCTTCAAGAGGACCATCCAGGGAGCATCGAGIACAGCTGTCCC 508

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||| |||||   |||   ||      | ||||||  ||| ||||||||||  |
Db        509 GCCTCCAATGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGTCAGGCCTGCCGCTTC 568

Qy        460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGCAGAGAGAGATGTGGGTAC 519
               |||||       |||||  ||| | |||   ||    ||    ||     |||    |
Db        569 ACCAAGTGCCTGCGGTGGGCATGCTCAAGGAGGCTGTGCGTCTGGA---CCGTGTCCGC 625

Qy        520 CGCCTTGTGCGGCAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCAAG 579
               ||     || ||   | | | |       |||  |   |  |   |   |  || |
Db        626 GGCGGACGGCAGAAGTACAAACGGCGGCCAGAGGTGGACGCTTTGCGTTTCGCGGGGCGC 685

Qy        580 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAG 639
                  || || |   | |   ||   | |  |    |  |  ||||       ||

Db        686 TTCCCTGCTGGACCTCTTGGCAGTAGCTGGAGGACCCAGGAAGACAGCCCCAGTGAACGCT 745

Qy        640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGCCCCAGT 699
              | | |   ||      ||  |||  |       |   |         ||      |||
Db        746 CTGGTGTCGCATCTGCTGGTGGTTGAACCTGAGAAGCTGTACGCCATGCCTGACCCAGCA 805

Qy        700 GCCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 759
                 |||       |   | ||    ||    | ||      |  |    ||    ||| |
Db        806 AGCCCCGATCGACACCTCCCCGCTGTGGCCACTCTCTGTGACCTTTTTGATCGAGAGATA 865

Qy        760 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
              ||   ||  |||||||||||||||  || || |||||       ||    |||| |||
Db        866 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCTCTCACTGTCACTGATCAG 925

Qy        820 CAAGTGCCGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGC 879
              ||  ||    |  || ||||   |||||||| ||| |  || ||||  ||      |||
Db        926 CACATGTCACTACTGCAGAGTGTGTGGATGCAAGTGCTGGTGCTGGGTGTGGCCCACCGC 985

Qy        880 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAG 939
              ||| |   |  |  |  ||||   |||||||   || || ||  |  |    || | |
Db        986 TCACTGCCACTGCAGGATGAGCTGGCCTTTGCTGAGGACCTGGTCTTATATGAAGAGGGG 1045

Qy        940 G 940
              |
Db       1046 G 1046


RESULT 20
BG747925
LOCUS       BG747925                 676 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602785778F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842402 5',
            mRNA sequence.
ACCESSION   BG747925
VERSION     BG747925.1  GI:14058578
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph D
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1676  row: 1  column: 19
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1..676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842402"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACCAC(G)   Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      113 a    209 c    222 g    142 t
ORIGIN
```

```
  Query Match             8.4%;  Score 106.2;  DB 12;  Length 676;
  Best Local Similarity   48.7%;  Pred. No. 2.5e-19;
  Matches  288;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;

Qy        280 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
              | | |||||   |||||  | ||    || || || || |||||||| ||    || |||
Db         26 CGCCTCTGCCTGGTCTGTGGGCACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 85

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || |||||| || || || | ||| || ||    |  || || |  ||||||
Db         86 GAGGCCTGCAAAGCCTTCTTCAAGAGCACCATCCAGGGAGCATCGAGTACAGCTGTCCG 145

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||  |||||   |||   ||     | ||||||  ||||||||||||||||  |
Db        146 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 205

Qy        460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTAC 519
                |||||       ||||| ||| | |||   ||    ||    ||  |     ||
Db        206 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGGCCTGGACCGCGTCCGGGGT 265

Qy        520 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 579
               | |    |  |   || | | | | | | |  |     |         ||
Db        266 CGGCGGCAGAAGTACAACGGCGGCCCGGAGGTGCACCCACTGCCCTTCCCGGGCCCCTTC 325

Qy        580 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCGAG 639
                 |||   ||  ||  |  |   |  | | |  |  ||||         |
Db        326 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCA 385

Qy        640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 699
              | | |    |      ||  |||  |       |   |          ||      ||||
Db        386 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCA 445

Qy        700 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 759
              |  ||         |   | |      ||    ||||     | |    |||   ||| |
Db        446 GGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGACATT 505

Qy        760 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
              ||   ||  ||||||||||||||||  || || |||||      |||    ||||  |||
Db        506 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 565

Qy        820 CAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTG 870
              ||  ||    |  || ||||||   ||||||||||||| |  || ||||  ||
Db        566 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTG 616


RESULT 21
BI100868
LOCUS       BI100868                 844 bp    mRNA    linear   EST 26 JUN 2001
DEFINITION  602886184F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5041580
            5', mRNA sequence.
ACCESSION   BI100868
VERSION     BI100868.1  GI:14551761
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11114  row: q  column: 21
            High quality sequence stop: 775.
FEATURES    .           Location/Qualifiers
     source          1. .844
                     /organism "Mus musculus"
                     /strain "FVB/N"
                     /db_xref "taxon:10090"
                     /clone="IMAGE:5041580"
                     /clone_lib "NCI_CGAP_Kid14"
                     /lab_host "DH10B (T1 phage resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      156 a    244 c    276 g    168 t
ORIGIN

  Query Match             8.1%;  Score 102;  DB 13;  Length 844;
  Best Local Similarity   48.7%;  Pred. No. 4.6e-18;
  Matches  307;  Conservative    0;  Mismatches  320;  Indels    3;  Gaps    1;

Qy        311 CGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCA 370
              |||| || ||||| || ||    || ||||| |  || || |||| || || || | ||
Db         11 CGGGCTACCACTACGGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGCACCA 70

Qy        371 TTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACC 430
              | || ||    |  || || |  ||||| ||  | ||| ||||   |||   ||
Db         71 TCCAGGGGAGCATCGAGTACAGCTGTCCGGCCTCCAATGAGTGTGAGATCACCAAGCGGA 130

Qy        431 GGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGT 490
              | ||||||  ||| ||||||||||  |    |||||       ||||| ||| | |||
Db        131 GACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAAGTGCCTGCGGGTGGGCATGCTCAAGG 190

Qy        491 GTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACG 550
                ||   ||   ||  |   |  | | |   |  |   |  | |  | ||| |
Db        191 AGGGTGTGCGTCTGGACCGTGTCCGCGGCGGACGGCAGAAGTACAAACGGCGGCCAGAGG 250

Qy        551 AGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGG 610
              | | |      |  |     ||     |||   |  ||    |  |
Db        251 TGGACCCTTTGCCTTTCCCGGGCCCCTTCCCTGCTGGACCTCTGGCAGTAGCTGGAGGAC 310

Qy        611 AGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGC 670
                |  |  | | |  ||| | |   |  |     ||      |  | ||  |||||
Db        311 CCAGGAAGACAGCCCCAGTGAACGCTCTGGTGTCGCATCTGCTGGTGGTTGAACCTGAGA 370

Qy        671 CGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGT 730
               ||   | |   ||      ||   | ||    |||        |   | ||    ||
Db        371 AGCTGTACGCCATGCCTGACCCAGCAAG---CCCCGATGGACACCTCCCCGCTGTGGCCA 427

Qy        731 CCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTC 790
              | ||     |  |     ||    ||| | ||   ||  ||||||||||||||  || |
Db        428 CTCTCTGTGACCTTTTTGATCGAGAGATAGTGGTCACCATCAGCTGGGCCAAGAGCATCC 487

Qy        791 CCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGG 850
              | |||||        ||    |||| |||||  ||    |  || ||||    |||||||
Db        488 CAGGCTTCTCCTCACTGTCACTGTCTGACCAGATGTCAGTACTGCAGAGTGTGTGGATGG 547

Qy        851 AGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTG 910
              | ||| |  || ||||  ||     |||||||  |    |  |  |  ||||   |||||
Db        548 AAGTGCTGGTGCTGGGTGTGGCCCAGCGCTCACTGCCACTGCAGGATGAGCTGGCCTTTG 607

Qy        911 CTCCAGATCTTGTTCTGGACAGGGATGAGG 940
              ||   || || || || ||  ||    || | ||
Db        608 CTGAGGACCTGGTCCTAGATGAAGAGGGGG 647


RESULT 22
BM562894
LOCUS       BM562894                 937 bp    mRNA    linear   EST 20-FEB 2002
DEFINITION  AGENCOURT_6588670 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478584
            5', mRNA sequence.
ACCESSION   BM562894
VERSION     BM562894.1  GI:18809360
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 947)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cqapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1998  row: h  column: 09
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1..947
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5478584"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTE7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed by Ying Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      182 a    286 c    297 g    170 t      2 others
ORIGIN

  Query Match             8.1%;  Score 101.4;  DB 13;  Length 937;
  Best Local Similarity   47.9%;  Pred. No. 7.3e-18;
  Matches  291;  Conservative    0;  Mismatches  316;  Indels    0;  Gaps    0;

Qy        334 TCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATT 393
              || ||||| |  || || ||||| || || || | ||| || ||    |  || || |
Db          5 TCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGC 64

Qy        394 TGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGC 453
              ||||| ||  | ||  |||||   |||   ||      | ||||||  ||||||||||||
Db         65 TGTCCGGCCTCCAACGACTCTGAGATCAGTAAGCGGACAGCAAGGCTGCCAGGCCTGC 124

Qy        454 CGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGT 513
              ||  |   |||||       ||||| ||| | |||   ||    ||    ||  |
Db        125 CGCTTCACCAAGTGCCTGCGGGTGGGCATGCTCAAGAGGGAGTGCGCCTGCAGCGGCTC 184

Qy        514 GGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCAGTGTGCCCCAAG 573
              ||    | |    | |    || | |  | || | | | |  |         |
Db        185 CGGGGTGGGCGGCAGAAGTACAAGCGGCGGCGGAGGTGGACCCACTGCCCTTCCCGGGC 244

Qy        574 GCCAACACAACTCGCCGCCACGGCCCCCACTGCCCCAGCTGCTGCTGCAGCCCTGACC 633
              ||         |||   ||  ||   |   |   |   |   |  |  |  ||||
Db        245 CCCTTCCCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGC 304

Qy        634 CCCGAGCAGCTAGTGCTCACCGTCCTGGAGGCTGAGCCGCCCCATGTCCTGATCAGCGAC 693
              |  | | |    |      ||  |||  |       |  |           ||      |
Db        305 AATGCACTGGTGTCTCATCTGCTGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTCAC 364

Qy        694 CCCAGTCGCGCCTTCACCGAGGCGTCCATCATGATGTCCCTGACCAAGTTGGCCGACAAG 753
              |||   |  ||      |  | |      ||   ||||     |  |    |||
Db        365 CCCGCAGGCCCTGATGGCCACCTCCCAGCCGTGGCTACCTCTGTGACCTCTTGACCGA 424

Qy        754 GAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGTTTGGTGGAGCTGACCGTC 813
              ||| | ||   ||  ||||||||||||||||  || || |||||      |||    |||
Db        425 GAGATTGTGGTCACCATCAGCTGGGCCAACAGCATCCCAGCCTTCTCATCGCTCTCGCTC 484

Qy        814 TTCGACCAAGTGCGGCTCTTGGACAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATG 873
```

```
              |  |||||  ||    |  || |||||  ||||||||||||  |  || ||||  ||
Db        485 TCTGACCAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCC 544

Qy        874 TGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTGCTCCAGATCTTGTTCTGGACAGG 933
              |||||||  |    |  |  |  ||||   ||| |||   ||  |  || |||||
Db        545 CAGCGCTCACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTAGTCCTGGATGAA 604

Qy        934 GATGAGG 940
              || | ||
Db        605 GAGGGGG 611


RESULT 23
BJ058831
LOCUS       BJ058831                 558 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ058831 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL060113 5', mRNA sequence.
ACCESSION   BJ058831
VERSION     BJ058831.1  GI:17492544
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia, Batrachia, Anura, Mesobatrachia, Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..558
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL060113"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
                     /note="Vector: pPSPN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett and A.M. Zorn,
                     (Wellcome/CRC Institute). "
BASE COUNT      142 a    134 c    146 g    136 t
ORIGIN

  Query Match             8.0%;  Score 100.4;  DB 13;  Length 558;
  Best Local Similarity   51.6%;  Pred. No. 1.1e-17;
  Matches  292;  Conservative    0;  Mismatches  256;  Indels   18;  Gaps    2;

Qy        307 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTAAAAGA 366
              ||||| || ||||||||||| ||    || ||||| | ||| || ||||| || ||  |
Db          1 GCATCAGGCTATCACTATGGGGTTGCTCATGTGAGGCATGCAAAGCCTTCTTCAAGCGC 60

Qy        367 AGCATTCAACGCACATAATGATTATATTTCTCCACCTACAAATCAGTGTACAATCCATAAA 426
              | ||| |||||  |||  || || |  |||||||| |  ||| | |||  |||    ||
Db         61 ACCATCCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATGAATGTGAAATTACCAAG 120

Qy        427 AACCGGCGCAAGAGCTGCCAGGCCTGCGACTTCGGAAGTGTTACGAAGTGGGAATGGTG 486
              |  || || |||  ||| || || |||||  |    |||||        ||||| ||| |
Db        121 AGACGTCGAAAGGCCTGTCAAGCTTGCCGCTTCACAAAGTGCCTTCGCGTGGGCATGCTA 180

Qy        487 AAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCC 546
              ||    ||   ||   ||  |    |    | | |  |       |||| ||  |
Db        181 AAACAAGCGCTACGCTTGGACCGTGTCAGAGCAGCACGTCAGAAATACAACAGGAGGCCA 240
```

```
Qy        547 GACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTG 606
              || |   | | | |                |       ||| || || || |  | ||
Db        241 GATGGAGATATCCTC------------CAGTACTCGTCTGGTGCCCACGCTCACCAAGCC 288

Qy        607 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 666
              |      || | |   | | || | | |  ||| || | | |||| || | |
Db        289 TCTGCAACAGTGGTCAAGAAGCAAACTCCAGTGAATGCAGTTGTCTCTCACCTGCTGGTT 348

Qy        667 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATG--- 723
              |    || |   | || |      ||     | | | | | || |       ||
Db        349 GCAGAACCAGACAAGTTGTTTGCTATGCCAGACCCAGCACTTCCTGATGGTGATTTGAAA 408

Qy        724 ---ATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
                 ||||  |||||    |  | || ||    ||  | ||   ||| || ||||||||
Db        409 TATATGAGCACCCTGTGTGACCTTGCTGATCGTGAAATTGTCATCATCATTAGCTGGGCA 468

Qy        781 AAGAAGATTCCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || || ||       ||   ||| |  ||||  || | |||||| ||||
Db        469 AAGAATATCCCAGGATTCTCCTCCCTTTCCCTCTCGGACCAGATGAGTCTCTTGCAGAGT 528

Qy        841 TGTTGGATGGAGGTGTTAATGATGGG 866
              ||||||||||| ||| || | ||
Db        529 GTCTGGATGGAGGTTTTATTGCTTGG 554


RESULT 24
BE546976
LOCUS       BE546976                 908 bp    mRNA    linear   EST 09-AUG-2000
DEFINITION  601071950F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3457885 5',
            mRNA sequence.
ACCESSION   BE546976
VERSION     BE546976.1  GI:9775621
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 908)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8448  row: d  column: 14
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1..908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3457885"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV SPORT6; Site_1: Not1;
                     Site_2: Sal1; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      172 a    251 c    344 g    151 t
ORIGIN

  Query Match             7.9%;  Score 99.8;  DB 10;  Length 908;
  Best Local Similarity   48.8%;  Pred. No. 2.1e-17;
  Matches  278;  Conservative    0;  Mismatches  297;  Indels    0;  Gaps    0;

Qy  1       897 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGA 956
```

```
                ||  ||  ||  ||  |||||||| ||     || ||||| |   || ||  ||||| ||  ||  ||
Db          11 GCCTCCGGCTACCACTATGGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGG 70

Qy        367 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTCTACAATCGATAAA 426
              | |||  ||  ||     |  || ||  |  ||||| ||  | ||  |||||   |||   ||
Db          71 ACCATCCAGGGGAGCATCGAGTACAGCTGTCCGGCCTCCAACGAGTGTGAGATCACCAAG 130

Qy        427 AACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTG 486
                | ||||||  ||||||||||||||||  |    |||||       ||||| ||| |
Db        131 CGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAAGTGCCTGCGGGTGGGCATGCTC 190

Qy        487 AAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCC 546
              |||   ||    ||    ||  |    ||    | |   | |    || |  | |  | |
Db        191 AAGGAGGGAGTGCGCCTGGACCCGCGTCCGGGGTGGGCGGCACAAGTACAACGGGCGGCG 250

Qy        547 GACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTG 606
              || | | | |  |         |       ||       ||| ||  || |   |  |
Db        251 GAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCCCTGCTGGGCCCCTGGCAGTCGCTGGA 310

Qy        607 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 666
              |    | |  |   |  ||||         | | | |   |    ||  |||  |
Db        311 GGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACTGGTGTCTCATCTGCTGGTGGTTGAG 370

Qy        667 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATG 726
              |   |         ||      ||||   |  ||        |  | |     ||
Db        371 CCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCCCTGATGGGCACCTCCCAGCCGTG 430

Qy        727 ATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAG 786
              ||||    |  |    |||  ||| | ||   ||  ||||||||||||||||
Db        431 GCTACCCTCTGTGACCTCTTTGACCGAGAGATTGTGGTCACCATCAGCTGGGCCAAGAGC 490

Qy        787 ATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGG 846
              || || |||||      |||    ||||  |||||  ||    |  || |||||   |||
Db        491 ATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAGATGTCAGTACTGCAGAGCGTGTGG 550

Qy        847 ATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTC 881
              |||||||| |  || |||  ||    ||| | |
Db        551 ATGGAGGTGCTGGTGCTGGGTGTGGCCAGGCTCAC 585


RESULT 25
AU169990
LOCUS       AU169990                 800 bp    mRNA    linear   EST 29-JAN-2001
DEFINITION  AU169990 Ol-br-ad cDNA Oryzias latipes cDNA clone br5464, mRNA
            sequence.
ACCESSION   AU169990
VERSION     AU169990.1  GI:12592059
KEYWORDS    EST.
SOURCE      Japanese medaka.
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryzinae; Oryzias.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Mita,K., Ishikawa,Y. and Yamauchi,M.
  TITLE     Establishment of cDNA database of medaka, Oryzias latipes
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mita K
            Genome Research Group
            National Institute of Radiological Sciences
            Anagawa 4-9 1, Inage, Chiba 263-8555, Japan
            Email: kmita@nirs.go.jp
            method:uni-directional sequence direction:sequenced from T4 primer
            (5' -> 3').
FEATURES             Location/Qualifiers
     source          1..800
                     /organism="Oryzias latipes"
                     /strain="HNI"
                     /db_xref="taxon:8090"
                     /clone="br5464"
                     /clone_lib="Ol-br-ad cDNA"
                     /sex="female/male mixed"
```

```
                     /tissue_type="brain"
                     /dev_stage="adult"
BASE COUNT       198 a    212 c    238 g    152 t
ORIGIN

  Query Match             7.9%;  Score 99.6;  DB 9;  Length 800;
  Best Local Similarity   65.0%;  Pred. No. 2.2e-17;
  Matches  147;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        284 TCTGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 343
              |||||   || ||  | |||   || || || ||||||||||||||   ||| |||||||
Db        568 TCTGCCTCGTATGTAGGGATGTGGCGTCCGGGTATCACTATGGAGTCGCCTCATCTCAAG 627

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              || || || || || || |||| |||||||||  | |  || || |  |||||||
Db        628 CCTGCAAAGCTTTCTTCAAGAGAACCATTCAAGGTAACATCGAGTACAGCTGTCCAGCAT 687

Qy        404 CAAATCAGGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              ||||  ||||   ||||   ||||  ||  | |||   ||||| || |||||  ||   |
Db        688 CAAACGAGTGCGAAATCACCAAAAGACGCAGAAGCCGTGCCAAGCATCCCGCTTTACCA 747

Qy        464 AGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAG 509
              ||||   | |||| || ||| | |||   ||   ||   ||| ||
Db        748 AGTGCCTCAAAGTAGGCATGCTCAAGGAGGGAGTGCGTCTACATAG 793


RESULT 26
AL527012
LOCUS       AL527012                1004 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL527012 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC017YG10 5
            prime, mRNA sequence.
ACCESSION   AL527012
VERSION     AL527012.1  GI:12790505
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .1004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DC017YG10"
                     /clone_lib="LTI_NFL003_NBC3"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed
                     by Life Technologies. Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT       163 a    324 c    350 g    164 t      3 others
ORIGIN

  Query Match             7.9%;  Score 99.6;  DB 9;  Length 1004;
  Best Local Similarity   49.0%;  Pred. No. 2.5e-17;
  Matches  290;  Conservative    1;  Mismatches  300;  Indels    1;  Gaps    1;

Qy        280 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGCTCGTGT 339

              | | |||||   |||||  | ||    || || || || |||||||| ||    || |||
Db        380 CGCCTCTGCCTGGTCTGTGGGCACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 439

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || |||||  || || | ||| || ||    |  || || |  ||||||
Db        440 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCA 499

Qy        400 GCTACAAATCAGGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||  ||||  |||   ||      | ||||||  |||||||||||||||  |
Db        500 GCCTCCAACGAGTGTGAAATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 559

Qy        460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGTAC 519
                |||||       ||||| ||| | |||   ||    ||   ||  |    ||
Db        560 ACCAAGTGCCTGGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGC 619

Qy        520 CGCCTGGGCGGAGACAGAGAAGTGCCACGAGCAGCTGCACTGGGCGGCAAGGCCAAG 579
                |   |       |||       |||| |   |   ||||| |   |  ||||
Db        620 TGGGGGCAGAAGTACAAGGGGCGCCGGAGGTGGGACCACTGCCCTTCGGGGCGCCT 679

Qy        580 ACAAGTGGCGGCCACGGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCGAG 639
                |||| |  ||       ||  ||  |       |   |          |
Db        680 TCCCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATG 739

Qy        640 CAGCTAGTGCCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 699
              ||       |||| |  |  |  |     | |              ||   |||   ||
Db        740 CACTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCCTATGCCATGCCTGACCCCG 799

Qy        700 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAACTTGGCCGACAAGGAGTTG 759
              |||||     |   | |   ||  ||||     |  |   |||  ||| |
Db        800 CAGCCCTGATGGGCACTCCCAGCCGTGCCTACCCTCTGTGACCTCTTTGACCCAGAGATT 859

Qy        760 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTCGTGGAGCTCAGCCTGTTCGAC 819
              ||   ||  ||||||||||||||||  || |  |||||      |||    ||||  |||
Db        860 GTGGTCACCATCAGCTGGGCCAAGAGCATCCAAGGCTTCTCATCGCTGTCGCTGTCTGAC 919

Qy        820 CAAGTGGGGCTCTTGGAGAGC-TGTGGATGGAGGGTATAATGATGGGGCTG 870
              ||  ||    |  || |||:| || ||||| |||||| |  || ||| | ||
Db        920 CAGATGTCAGTACTGCAGAPCGTGGTGGATTGAGGTGCTGGTGCTGGTGGTG 971


RESULT 27
BM933826/c
LOCUS       BM933826                528 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  UI-M-BH3-bsc-b-09-0-UI.r2 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-bsc-b-09-0-UI 5', mRNA sequence.
ACCESSION   BM933826
VERSION     BM933826.1  GI:19392978
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .528
```

/organism "Mus musculus"
/strain "C57BL/6J"
/db_xref "taxon:10090"
/clone "UI-M-BH4-bse-b-09-0-UI"
/clone_lib "NIH_BMAP_M_S4"
/dev_stage "27-42 days"
/lab_host "DH10B (LifeTechnologies)"
/note "Vector: pIVT4D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Ronaldo, Tennor and Soares, Genome Research 6:791-806, 1996)"
BASE COUNT 119 a 124 c 156 g 128 t 2 others
ORIGIN

Query Match 7.8%; Score 97.8; DB 14; Length 528;
Best Local Similarity 57.8%; Pred. No. 5.8e-17;
Matches 193; Conservative 0; Mismatches 138; Indels 3; Gaps 1;

```
Qy        917 ATCTTGTTCTGGACAGGGATGAGGGGAAATGCCTAGAAGGAATTCTGGAAATCTTTGACA 976
              |  |   | ||||||||| || | || ||||| || ||||| ||  |||| ||||||||||
Db        525 ACNTGCTCCTGGACAGGAATCAAGGTAAATGTGTGGAAGGCATGGTGGAGATCTTTGACA 466

Qy        977 TGCTCCTGGCAACTACTTCAAGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCT 1036
              || | ||||| ||  |     |||| ||   | | || || ||     |||| | | | |
Db        465 TGTTGCTGGCTACGTCAAGTCGGTTCCGCATGATGAACCTGCAGGGAGAAGAGTTTGTGT 406

Qy       1037 GTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCC--TCTGGTCACAGCGACCC 1093
              |  |||| |||| ||  |||| |||||| |  ||||| |   |||| ||   |
Db        405 GCCTCAAATCCATCATTTTGCTTAATTCCGGAGTGTACACGTTTCTGTCCAGCACCTTGA 346

Qy       1094 AGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGG 1153
              ||  |   ||     |   |  |    | ||| ||    | || ||  |||||
Db        345 AGTCTCTGGAAGAGAAGGACCACATCCACCGTGTCCTGGACAAGATCACAGACACTTTGA 286

Qy       1154 TTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACC 1213
              |    |||| |||||    ||| |  |   ||||||||        ||||| || | |
Db        285 TCCACCTGATGGCCAAAGCTGGCCTGACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGC 226

Qy       1214 TCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              ||||  |  | || |||||  || ||||   |||
Db        225 TCCTTCTCATTCTTTCCCATATCCGGCACATGAG 192
```

RESULT 28
BG480960
LOCUS BG480960 775 bp mRNA linear EST 21-MAR-2001
DEFINITION 602529118F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4652501 5', mRNA sequence.
ACCESSION BG480960
VERSION BG480960.1 GI:13413239
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 775)
AUTHORS Mgc Project http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cqapbs r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1437 row: d column: 06
High quality sequence stop: 664.
FEATURES Location/Qualifiers
source 1..775
/organism "Homo sapiens"
/db_xref "taxon:9606"
/clone "IMAGE:4652501"
/clone_lib "NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note "Organ: placenta; Vector: pcDNA7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 181 a 216 c 282 g 146 t
ORIGIN

Query Match 7.7%; Score 96.6; DB 12; Length 775;
Best Local Similarity 51.1%; Pred. No. 1.6e-16;
Matches 303; Conservative 0; Mismatches 284; Indels 6; Gaps 3;

```
Qy        280 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
              | | |||||  |||||  | ||    || || || || ||||||| ||    || |||
Db         83 CGCCICTGCCTGGTCTGTGGGGACGTGGCTCCGGCTACCACTATGGTGTGGCATCTGT 142

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || ||||| || || || | ||| || |||   |  || || | |||||
Db        143 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGA-GCATCGAGTACAGCTGTCCG 201

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||  |||||   |||    ||     | ||||||  ||||||||||||||  |
Db        202 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 261

Qy        460 CGGAAGTGTTACGAAGTGGGAATGGTGAAG TGTGGCTCCCGGAGAGAGAGATGTGGGT 517
                |||||       ||||| ||| | |||   |  |  | |   ||  | |  | ||
Db        262 ACCAAGTGCCTGCGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCGGGGTG 321

Qy        518 ACCGCCTTGTGCCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCCGCAAGGCCA 577
              || |    |  | | | |    | |  || |  |||| ||   | |||     |
Db        322 GGCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCC 381

Qy        578 AGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCG 637
              |  |  |  |  | ||| |  | |  |  |  |||  ||  |     |
Db        382 CTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAAIGCAC 441

Qy        638 AGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCA 697
              |  |       ||     | | ||| ||||| || | |||   ||     || | |
Db        442 TGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAG 501

Qy        698 GTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGT 757
              |  |         |  | || ||  ||||    | |  |||   |||
Db        502 GCCCTGATGGGCACCTCCCAGCCGTGGCTA---CCCTCTGTGACCTCTTTGACGGAGAG 558
```

```
Qy        758 TGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCG 817
              | ||   ||  ||||||||||||||||  || || ||||       |||    ||||  |
Db        559 TTGTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTGGGTGTCTG 618

Qy        818 ACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTG 870
              ||||  ||    |  || |||||   ||||||||||||| |  || ||||  ||
Db        619 ACCAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTG 671


RESULT 29
BG033384
LOCUS       BG033384                979 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  602401556F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4403197 5',
            mRNA sequence.
ACCESSION   BG033384
VERSION     BG033384.1  GI:12425627
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10112  row: p  column: 14
            High quality sequence stop: 681.
FEATURES             Location/Qualifiers
     source          1. .979
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4403197"
                     /clone_lib="NIH_MGC_87"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      150 a    291 c    353 g    184 t      1 others
ORIGIN

  Query Match             7.7%;  Score 96.4;  DB 12;  Length 979;
  Best Local Similarity   49.0%;  Pred. No. 2.1e-16;
  Matches  285;  Conservative    0;  Mismatches  296;  Indels    1;  Gaps    1;

Qy        307 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGCATGTAAGGCCTTTTTTAAAAGA 366
              || || || || |||||||| ||    || |||| | || || |||| || || ||
Db          4 GCCTCCGGCTACCACTATGGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGA 63

Qy        367 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCCATAAA 426
              | ||| || |||   |  || || |  |||||||  | ||  |||||   |||   ||
Db         64 ACCATCCAGGGAGC-ATCGAGTACAGCTGTCCAGCCTCCAACGACTCTGACATCACCAAC 122

Qy        427 AACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACCAAGTGCGAATGCTG 486
              | ||||||  |||||||||||||||  |    |||||       ||||| ||| |
Db        123 CGGAGACGCAAGGCCTGCAGGCCTGCCGCTTCACCAAGTGCCTGCGGCTCGGCATGCTG 182

Qy        487 AAGTCTGGCTCCCGGAGAGAGAGATGTGGGTACCGCTTCTGCGCCAGACACACAAGTGCC 546
              |||   ||    ||    ||  |    ||    | |    |  |    || |  |  |
Db        183 AAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGTGGGCGGCAGAAGTACAAGCGGCGGCCG 242

Qy        547 GACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTG 606
              || | | | |  |       |    ||       |||   ||  ||   |  |
Db        243 GAGGTGGACCCACTGCCCTTCCCGGGCTGCTTCCCTGCTGGGCCCCTGGCAGTCGCTGGA 302

Qy        607 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 666
              |    | |  |   |  ||||        |  | | |    |     ||  |||  |
Db        303 GGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACTGCTGTCTCATCTGCTGGTGGTTGAG 362

Qy        667 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTCACCGAGGCCTCCATGATG 726
              |   |         ||      ||||  |  ||        |  | |     ||
Db        363 CCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCCCTGATGGGCACCTCCCAGCCGTG 422

Qy        727 ATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAG 786
              ||||     |  |    |||  ||| | ||   ||  ||||||||||||||||
Db        423 CCTACCCTCTGTGACCTCTTTGACCGAGAGATTGTGCTCACCATCAGCTGGGCCAAGAGC 482

Qy        787 ATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTCGAGAGCTGTTGG 846
              || || |||||     |||   |||| |||||  ||    |  || |||||    |||
Db        483 ATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAGATGTCAGTACTGCAGAGCGTGTGG 542

Qy        847 ATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGAC 888
              |||||||||| |  || ||||  ||     ||||||| |  ||
Db        543 ATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGCTCACTGCAC 584


RESULT 30
BI738331
LOCUS       BI738331                818 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  602361587F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5368892 5',
            mRNA sequence.
ACCESSION   BI738331
VERSION     BI738331.1  GI:15715344
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11938  row: m  column: 21
            High quality sequence stop: 817.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5368892"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      185 a    245 c    230 g    158 t
ORIGIN

  Query Match             7.6%;  Score 95;  DB 13;  Length 818;
  Best Local Similarity   65.1%;  Pred. No. 4.8e-16;
  Matches  140;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        280 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
```

```
                | | | |||   || ||| | ||    || || || || ||||| |||||    || ||
Db          303 CGCCTGTGCCTCGTGTGCGGGGACATTGCCTCTGGCTACCACTACGGAGTGGCCTCCTGC 362

Qy          340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
                || |  || ||||| || || || || |||| |||||||||  | |  || || |  || ||
Db          363 GAGGCTTGCAAGGCGTTCTTCAAGAGAACCATTCAAGGCAACATCGAGTACAACTGCCCG 422

Qy          400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCCACTT 459
                || || ||| | |||   |||   |||    |||||||| ||| ||||||||||||| |
Db          423 GCCACCAATGAATGTGAGATCACCAAACGGAGGCGCAAGTCCTGTCAGGCCTGCCGATTC 482

Qy          460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 494
                ||| ||   | |||||||| ||| |||||   ||
Db          483 ATGAAATGCCTCAAAGTGGGGATGCTGAAGGAAGG 517


RESULT 41
BM957249
LOCUS       BM957249                669 bp    mRNA    linear   EST 18-MAR-2002
DEFINITION  ty75h10.y1 Suqano SJD adult male Danio rerio cDNA clone 5777131 5'
            similar to TR:Q90272 Q90272 PETINOIC ACID RECEPTOR ALPHA 2.B. ;,
            mRNA sequence.
ACCESSION   BM957249
VERSION     BM957249.1  GI:19540679
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopteryqii; Neopteryqii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,R.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: ty75h10.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Suqano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I M A G E
            Consortium/LLNL, send email to: info@image.llnl.gov
            High quality sequence stop: 526.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5777131"
                     /clone_lib="Suqano SJD adult male"
                     /sex="male"
                     /tissue_type="whole body"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: BraI1 (CACCATGTG);
                     Site_2: BraI1 (CACTGTGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  BraI1 adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct BraI1
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed and donated by Dr. Sumio
                     Suqano (University of Tokyo Institute of Medical Science).
                     Custom primers for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
BASE COUNT      182 a    174 c    183 q    130 t
ORIGIN

  Query Match             7.3%;  Score 91.8;  DB 14;  Length 669;
  Best Local Similarity   60.7%;  Pred. No. 4.6e-15;
  Matches  150;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy          242 CCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCG 301
                |||||||| | |    ||| | ||    |||       |   |||   |||||||   |
Db          220 CCAGCCCTCCTTCACCGCCACCTCCACCACGGGTCTACAAGCCCTGTTTTGTCTGCCAGG 279

Qy          302 ATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTA 361
                |  |  | ||||| || ||||||||||||| |  | ||||||||  |||||||  |||||
Db          280 ACAAGTCTTCGGGCTACCACTATGGAGTCAGCGCATGTGAAGGGTCTAAGGGTTTTTTCC 339

Qy          362 AAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCG 421
                 ||||||||||||      | |  |  || |  || |         ||    ||||  |||
Db          340 GAAGAAGCATTCAGAAGAACATGGTGTACACCTGCCACAGAGAGAAGAGCTGTATCATCA 399

Qy          422 ATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCCACTTCGGAAGTGTTACGAAGTGGGAA 481
                | ||    |  || ||  ||||||||  ||||||||| | ||||||    |||||||| |
Db          400 ACAAGGTCACCCGAAACCGCTGCCAGTACTGCCGACTGCAGAAGTGCCTGGAAGTGGGCA 459

Qy          482 TGGTGAA 488
                ||   ||
Db          460 TGTCTAA 466


RESULT 42
BB711891
LOCUS       BB711891                434 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  BB711891 RIKEN full length enriched, 2 cells egg Mus musculus cDNA
            clone B020030L17 3', mRNA sequence.
ACCESSION   BB711891
VERSION     BB711891.1  GI:16065060
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanaqaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Soqabe,Y., Suzuki,H., Taqawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanaqawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: qenome-res@qsc.riken.qo.jp,
            URL: http://qenome.qsc.riken.qo.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Suqahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            qenes. Genome Res. . 10 (10), 1617-1630 (2000)
            waqi,K., Fujiwake,S., Inoue,K., Toqawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
```

```
           10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
           ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
           encyclopedia: real-time sequence clustering for construction of a
           nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
           further details.
            e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Mus musculus"
                     /db_xref "taxon:10090"
                     /clone="B020030L17"
                     /clone_lib="RIKEN full-length enriched, 2 cells egg"
                     /tissue_type="egg"
                     /dev_stage="2 cells"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGCATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGACTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I "
BASE COUNT       116 a     99 c    107 g    112 t
ORIGIN

  Query Match             7.3%;  Score 91.2;  DB 10;  Length 434;
  Best Local Similarity   66.0%;  Pred. No. 4.2e-15;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTCGTGAAG 343
              ||||||   || |||  |||| | || |||||||  || ||||| ||    | |||||| |
Db        219 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGTGCTGACGTGTGCAA 278

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| | ||| ||||||||     | |||||||  |  ||||  ||||| | |  |
Db        279 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 338

Qy        404 CAAATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | || |  || |||||||  ||  | || | ||| |  || |||||   ||| |
Db        339 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 398

Qy        464 AGTGTTACGAAGTGGGAATG 483
              | |||    |||  ||||||
Db        399 AATGTCTTCAAGCTGGAATG 418


RESULT 33
BG687711
LOCUS       BG687711                 521 bp    mRNA    linear   EST 02-MAY-2001
DEFINITION  334912 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BG687711
VERSION     BG687711.1  GI:13029512
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and
            Wells,K.D.
  TITLE     Mapping of Expressed Sequence Tags from a normalized Bovine mammary
            gland cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@anri.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 33  row: E  column: 5
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="BARC 5BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."
BASE COUNT       162 a    111 c    114 g    134 t
ORIGIN

  Query Match             7.3%;  Score 91.2;  DB 12;  Length 521;
  Best Local Similarity   66.0%;  Pred. No. 4.6e-15;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTCGTGAAG 343
              ||||||   || |||  |||| | || |||||||  || ||||| ||    | |||||| |
Db        225 TCTGCCTAGTCTGCTCTGATGAAGCTTCAGGATGTCATTATGGGTCTTGACTTGTGGAT 284

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| | ||| ||||||||     | |||||||  |  ||||| ||||| | |  |
Db        285 GCTGTAAAGTATTCTTCAAAAGAGCAGTGGAAGGACAGCACAATTATCTTTGTGCTGGAA 344

Qy        404 CAAATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | ||||  || |||||||  ||  | || | ||||  || |||||   ||| |
Db        345 GAAATGATTGTATCATTGATAAAATTCGAAGAAAAAACTGCCCAGCATGCCGCTATAGAA 404

Qy        464 AGTGTTACGAAGTGGGAATG 483
              | ||    |||  ||||||
Db        405 AATGCCTTCAAGCTGGAATG 424


RESULT 34
BB196002
LOCUS       BB196002                 698 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB196002 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330092E23 3' similar to X13358 Mouse mRNA for
            glucocorticoid receptor form A, mRNA sequence.
ACCESSION   BB196002
VERSION     BB196002.2  GI:15409595
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,I., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
```

```
  JOURNAL   Unpublished (2001)
COMMENT     On Jun 30, 2000 this sequence version replaced gi:8860955.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7 22 Suehiro-cho, Tsurumi ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81 45 503-9222
            Fax: 81 45 503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             Waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN,
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .698
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone "A330092E23"
                     /clone_lib "RIKEN full length enriched, adult male spinal
                     cord"
                     /sex "male"
                     /tissue_type "spinal cord"
                     /dev_stage "adult"
                     /lab_host "DH10B"
                     /note "Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo activated reverse-
                     transcriptase and subsequently enriched for full length by
                     cap trapper. cDNA went through one round of normalization
                     to Rot    10.0 and subtraction to Rot    459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT       193 a    149 c    173 g    184 t
ORIGIN

  Query Match             7.3%;  Score 91.2;  DB 10;  Length 698;
  Best Local Similarity   66.0%;  Pred. No. 5.5e-15;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 343
              ||||||   || |||  |||| | || |||||||  || ||||| ||   | ||||||
Db        485 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGGTGCTGACGTGTGCAA 544

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| | ||| |||||||||     |  |||||||  |  ||||  ||||| | |  |
Db        545 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 604

Qy        404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | || |  || |||||||  ||  | || | ||| |  || |||||   ||| |
Db        605 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAACTGTCCAGCATGCCGCTATCGAA 664

Qy        464 AGTGTTACGAAGTGGGAATG 483
              | |||   ||| ||||||
Db        665 AATGTCTTCAAGCTGGAATG 684


RESULT 35
BC016462
LOCUS       BC016462                1638 bp    mRNA    linear   HTC 07-AUG-2002
DEFINITION  Mus musculus, Similar to nuclear receptor subfamily 4, group C,
            member 1, clone IMAGE:4036433, mRNA.
ACCESSION   BC016462
VERSION     BC016462.1  GI:16741246
KEYWORDS    HTC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1638)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amq@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, K.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 24 Row: c Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6680102
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1. .1638
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /map="C57BL/6J"
                     /clone "IMAGE:4036433"
                     /tissue_type="Mammary tumor, WAP-TGF alpha model, 7 months
                     old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam5"
                     /lab_host "DH10B"
                     /note "Vector: pCMV-SPORT6"
BASE COUNT       446 a    396 c    433 g    363 t
ORIGIN
```

```
  Query Match             7.3%;  Score 91.2;  DB 11;  Length 1638;
  Best Local Similarity   66.0%;  Pred. No. 8.9e-15;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGCTCTGCAGCGATTACGCATCGGGATATCACTATCGAGTCTGGTCGTCGTCAAC 343
              |||||   || |||  |||| | || |||||||  || ||||| ||   | ||||||  |
Db       1403 TCTGCCTGCTGTGCTCCGATCAAGCTTCGGGATGCCATTATGGGGTGCTGACCTCTCCAA 1462

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGCACATAATGATTATATTTGTCCAGCTA 403
              | |||||  | ||| |||||||||     |  |||||||  |  ||||  ||||| | |  |
Db       1463 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGCAGCAGCAATTACCTTTGTGCTGCAA 1522

Qy        404 CAAATCAGTGTACAATCGATAAAAACGGGCGCAACAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | || |  || |||||||  ||  | || | ||| |  || |||||   ||| |
Db       1523 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 1582

Qy        464 AGTGTTACGAAGTGCGAATG 483
              | |||    |||  ||||||
Db       1583 AATGTCTTCAAGCTGGAATG 1602


RESULT 36
BG010872
LOCUS       BG010872                399 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  PC5-GN0194-061200-024-F11 GN0194 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG010872
VERSION     BG010872.1  GI:12458558
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the PAFESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getbtml2.pl?il%5Ft2_pc5_GN0194_
            061200_024_F11&t3_2000_12_06&t4_1)
            Seq primer: puc 18 forward
            High quality sequence stop: 70.
FEATURES             Location/Qualifiers
     source          1..399
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0194"
                     /dev_stage="Adult"
                     /note "Organ: placenta_normal; Vector: puc18; Site_1: Smal
                     ; Site_2: Smal; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      107 a     86 c    111 g     95 t
ORIGIN

  Query Match             7.2%;  Score 91;  DB 12;  Length 399;
  Best Local Similarity   63.5%;  Pred. No. 4.5e-15;
  Matches  139;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        292 GTCTGCAGCGATTACGCATCGGGATATCACTATCGAGTCTGGTCGTCGTCGAAGGATGTAAG 351
              || ||  | ||    ||| |  || || |||||||| ||    || |||||||  || |||
Db        148 GTGTGTGGTGACATCGCTTTTCGGTACCACTATGGGGTAGCAACATCTCAAGCCTGCAAC 207

Qy        352 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 411
              || || || || || |  ||||||||  |||  || || |  || || || || ||| |
Db        208 GCATTCTTCAACAGGACAATTCAAGGCAATATAGAATACAGCTGCCCTGCCAGCAATGAA 267

Qy        412 TGTACAATCGATAAAAACGGGCGCAACAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTAC 471
              |||  ||||   ||   | | || ||   |||||||| ||||  |   ||||||||
Db        268 TGTGAAATCACAAAGCGCAGACGTAAATCCTGCCAGGCTTGCCGCTTCATGAAGTGTTTA 327

Qy        472 GAAGTGCGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||||| ||| ||||    ||    ||    || |||
Db        328 AAACTGGGCATGCTGAAAGAAGCGGTCGCTCTTGACAGA 366


RESULT 37
BG746447
LOCUS       BG746447                592 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602703728F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4856886 5',
            mRNA sequence.
ACCESSION   BG746447
VERSION     BG746447.1  GI:14057100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I M A G E  Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1708  row: h  column: 07
            High quality sequence stop: 587.
FEATURES             Location/Qualifiers
     source          1..592
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:4856886"
                     /clone_lib "NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host "DH10B (phage resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      98 a    182 c    197 g    115 t
ORIGIN

  Query Match             7.2%;  Score 90.6;  DB 12;  Length 592;
  Best Local Similarity   48.5%;  Pred. No. 7.4e-15;
  Matches  278;  Conservative    0;  Mismatches  294;  Indels    1;  Gaps    1;

Qy        285 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATCGAGTCTGGTCGTCGTCGAAGG 344
              ||||  |||||  | ||    || || || || ||||||| ||    || ||||| |
```

```
Db          2 CTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGTGAGGC 61

Qy        345 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 404
              || || ||||| || || || | ||| || |||   |  || || |  ||||| ||  |
Db         62 CTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGAGC-ATCGAGTACAGCTGTCCGGCCTC 120

Qy        405 AAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 464
              ||  ||||  ||| ||  | |||||| ||||||||||||||  |    ||
Db        121 CAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAA 180

Qy        465 GTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCT 524
              |||       ||||| ||| | |||   ||    ||    ||  |     ||    | |
Db        181 GTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGTGGGGG 240

Qy        525 TGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAG 584
              | |    || | |  | || | | | |  |       |       ||
Db        241 GCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCCCTGC 300

Qy        585 TGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGTGGACGCCCTGAGCCCCGAGCAGCT 644
              |||   ||  ||   |  |  |   | | |  |  ||||    |  | | | |
Db        301 TGGGCCCCTGGCAGTCGCTGGAGGCCCCGGAAGACAGCAGCCCCAGTGAATGCACTGGT 360

Qy        645 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCC 704
              |    ||  ||  |  |    |  |      ||     ||||   |  ||
Db        361 GTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCC 420

Qy        705 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACA 764
              |  | |    ||   ||||    | |    |||   ||| | ||
Db        421 TGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATTGTGGT 480

Qy        765 CATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGT 824
              ||  ||||||||||||||||  || || |||||      |||    ||||  |||||  |
Db        481 CACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAGAT 540

Qy        825 GCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTT 857
              |    |  || |||||   ||||||||||||  |
Db        541 GTCAGTACTGCAGACGCTGTGGATGGAGGTGCT 573


RESULT 48
AW114877/c
LOCUS       AW114877                413 bp    mRNA    linear   EST 31 JAN 2000
DEFINITION  MC809 mouse liver, vehicle control Mus musculus cDNA clone MC809 3'
            similar to X13358 Mouse mRNA for glucocorticoid receptor form A,
            mRNA sequence.
ACCESSION   AW114877
VERSION     AW114877.1  GI:6826590
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Thomas,R.S., Rank,D.R., Penn,S.G., Zastrow,G.M., Jovanovich,S. and
            Bradfield,C.A.
  TITLE     McArdle Laboratory/Molecular Dynamics Dioxin EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Bradfield CA
            McArdle Laboratory for Cancer Research
            University of Wisconsin
            1400 University Ave., Madison, WI 53706, USA
            Tel: 608 262 2024
            Fax: 608 262 2824
            Email: bradfield@oncology.wisc.edu
            This clone was sequenced as part of a project to develop a database
            on gene expression changes following exposure to various
            environmental toxicants. The database can be accessed at
            http://mcardle.oncology.wisc.edu/bradfield/  Treatment: dioxane
            vehicle (400 ul/kg). Animals sacrified: 48 hrs post-injection.
            Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1..413
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="MC809"
                     /clone_lib="mouse liver, vehicle control"
                     /sex "male"
                     /lab_host "JM109"
                     /note="Organ: liver; Vector: pGEM11Zf (Promega); Site_1:
                     NotI; Site_2: EcoRI; First strand cDNA was primed with a
                     NotI-polyT primer
                     [5'-AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT-3'].
                     Double-stranded cDNA was ligated with EcoRI adapters
                     (Pharmacia), digested with NotI, and ligated into the
                     EcoRI/NotI sites of the pGEM11Zf vector. The library was
                     NOT normalized."
BASE COUNT      104 a    105 c     98 g    104 t      4 others
ORIGIN

  Query Match             7.2%;  Score 90.2;  DB 10;  Length 413;
  Best Local Similarity   65.5%;  Pred. No. 7.9e-15;
  Matches  141;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 343
              |||||   || |||  |||| | || |||||||  || ||||| ||   | ||||||
Db        214 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGGTGCTGACGTGTGGAA 155

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| | ||| |||||||||    | |||| ||  |  |||| |||| | | |
Db        154 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGNCAGCACAATTACCTTTGTGCTGGAA 95

Qy        404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | || |  || ||||||  ||  | || | ||| |  || ||||   ||| |
Db         94 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 35

Qy        464 AGTGTTACGAAGTGGGAATG 483
              | |||    |||  ||||||
Db         34 AATGTCTTCAAGCTGGAATG 15


RESULT 49
AL547737
LOCUS       AL547737                1010 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL547737 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1017YH04 5
            prime, mRNA sequence.
ACCESSION   AL547737
VERSION     AL547737.1  GI:12882082
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1010)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..1010
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0D1017YH04"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note "Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
```

Email : tliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 171 a 303 c 349 g 182 t 5 others
ORIGIN

Query Match 7.2%; Score 90.2; DB 9; Length 1010;
Best Local Similarity 50.8%; Pred. No. 1.3e-14;
Matches 295; Conservative 0; Mismatches 278; Indels 8; Gaps 3;

Qy 280 CACTTCTCCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCCATCT 339
Db 293 CGCCTCTGCCTGGTCTGTCGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 352

Qy 340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTGTCCA 399
Db 353 GAGGCCTGCAAAGCCTTCTTCAAGACGACCATCCAGGGCAGCATCGAGTACAGCTGTCCG 412

Qy 400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
Db 413 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAACCCCTGCCAGCCCTCCCCCTTC 472

Qy 460 CGGAAGTGTTACGAAGTGGCAATGGTGAAGTGTGGCTCCCGCAGAGAGATGTGGGTAC 519
Db 473 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCCCCCT 532

Qy 520 CGCCTTGTGCGGGAGACAGACAAGTGCCGA--CGAGCAGCTGCACTGTGCGGGCAAGGCCA 577
Db 533 GGGCGGCAGAAGTACAAGCGGCGGCCGGAGTGCACCCACTGCCCTTCCCCGCCCCTTCC 592

Qy 578 AGAGAAGTGGCGGCCAGGGGCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTCAGCCCCG 637
Db 593 CTGCTGGGCCCTGGCAGTGCCTGGAGGCCCCGGAAGACAGCAGCCCCAGTGAATGCAG 652

Qy 638 AGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGCCCCCA 697
Db 653 TGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAG 712

Qy 698 GTGC-GCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAG 756
Db 713 GCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATTG 772

Qy 757 TTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTC 816
Db 773 TGGTCAC-----ATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTGGCTGTC 827

Qy 817 GACCAAGTGCCGCTCTTGGAGAGCTGTTGGATGCAGGTGTT 857
Db 828 GACCAGATGTCAGTACTGCAGAGCGTGTGGATTGAAGTGCT 868

RESULT 40
BB196485
LOCUS BB196485 275 bp mRNA linear EST 30-JUN-2000
DEFINITION BB196485 RIKEN full-length enriched, adult male spinal cord Mus musculus cDNA clone A330094K06 3' similar to X13358 Mouse mRNA for glucocorticoid receptor form A, mRNA sequence.
ACCESSION BB196485
VERSION BB196485.1 GI:8861438
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 275)
AUTHORS Kondo,H., Aizawa,K., Akahira,S., Akiyama,T., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,I., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,I., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,T., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1..275
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A330094K06"
/clone_lib="RIKEN full-length enriched, adult male spinal cord"
/sex="male"
/tissue_type="spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. "
BASE COUNT 75 a 65 c 68 g 67 t
ORIGIN

Query Match 7.1%; Score 89.6; DB 10; Length 275;
Best Local Similarity 64.4%; Pred. No. 9.8e-15;
Matches 134; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 276 TGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTC 335
Db 54 TCGCCAACTCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTGTGGGGTGCTGAC 113

Qy 336 GTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTGTC 395
Db 114 GTGCTGGAAGCTGTAAATTCTTCTTTAAAAGAGGCAGTGGAAGGACAGCACAATTACCTTG 173

Qy 396 TCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCG 455

```
                 | | |  | |||| | || |  || |||||||  ||  | || | ||| |  || |||||
Db        174 TGCTGGAAGAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCG 233

Qy        456 ACTTCGGAAGTGTTACGAAGTGGGAATG 483
                  ||| || |||    |||  ||||||
Db        234 CTATCGAAAATGTCTTCAAGCTGGAATG 261


Search completed: June 16, 2003, 04:16:08
Job time : 1914.3 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 00:23:52 ; Search time 5026.77 Seconds
(without alignments)
10988.581 Million cell updates/sec

Title: US-08-826-361A-24
Perfect score: 1898
Sequence: 1 CACGAATCTTTGAGAACATT.....TGAATCGAGCGTTTCCACAC 1898

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : GenEmbl:*
1: qb_ba:*
2: qb_htg:*
3: qb_in:*
4: qb_om:*
5: qb_ov:*
6: qb_pat:*
7: qb_ph:*
8: qb_pl:*
9: qb_pr:*
10: qb_ro:*
11: qb_sts:*
12: qb_sy:*
13: qb_un:*
14: qb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1898	100.0	1898	6	A65127	A65127 Sequence 24
2	1898	100.0	1898	6	AX323078	AX323078 Sequence
3	1718	90.5	1740	9	AB006590	AB006590 Homo sapi
4	1669	87.9	2011	6	AX234658	AX234658 Sequence
5	1669	87.9	2011	9	AF051427	AF051427 Homo sapi
6	1635.6	86.2	1647	6	AX029400	AX029400 Sequence
7	1558	82.1	1560	9	HSRNAFRR	X99101 H.sapiens m
8	1494.2	78.7	1651	9	CJERB	Y09372 C.jacchus m
9	1482	78.1	2041	9	AF051428	AF051428 Homo sapi
10	1482	78.1	2745	9	AF060555	AF060555 Homo sapi
11	1480.4	78.0	3593	9	AB006589	AB006589 Homo sapi
12	1453.6	76.6	1460	6	A61586	A61586 Sequence 4
13	1453.6	76.6	1460	6	AR075921	AR075921 Sequence
14	1453.6	76.6	1460	6	AX022458	AX022458 Sequence
15	1434	75.6	1434	6	A65104	A65104 Sequence 1
16	1434	75.6	1434	6	AX323055	AX323055 Sequence
17	1375.8	72.5	1640	9	AF393815	AF393815 Macaca ar
18	1308.2	68.9	1566	9	AF393816	AF393816 Callithri
19	1256.4	66.2	1820	4	BTA18017	Y18017 Bos taurus
20	1256.2	66.2	1611	4	AF110402	AF110402 Bos tauru
21	1254.6	66.1	1584	4	AF177936	AF177936 Ovis arie
22	1247	65.7	1251	6	A65105	A65105 Sequence 2
23	1247	65.7	1251	6	AX323056	AX323056 Sequence
24	1247	65.7	1257	6	A65123	A65123 Sequence 20
25	1247	65.7	1257	6	AX323074	AX323074 Sequence
26	1240.2	65.3	1650	10	RNAJ2602	AJ002602 Rattus no
27	1239.6	65.3	2555	10	RNU57439	U57439 Rattus norv
28	1239.6	65.3	2568	6	A61583	A61583 Sequence 1
29	1239.6	65.3	2568	6	AR075920	AR075920 Sequence
30	1239.6	65.3	2568	6	AX022455	AX022455 Sequence
31	1227.4	64.7	1581	4	AF164957	AF164957 Sus scrof
32	1227.4	64.7	1581	4	AF267736	AF267736 Sus scrof
33	1194	62.9	2153	9	BC024181	BC024181 Homo sapi
34	1169	61.6	1427	9	AF124790	AF124790 Homo sapi
35	1130	59.5	1621	10	AF042058	AF042058 Rattus no
36	1117.4	58.9	2152	10	MM081451	U81451 Mus musculu
37	1099.6	57.9	1458	6	A61588	A61588 Sequence 6
38	1099.6	57.9	1458	6	AR075922	AR075922 Sequence
39	1099.6	57.9	1458	6	AX022460	AX022460 Sequence
40	1099.6	57.9	1458	10	MMAJ220	AJ000220 Mus muscu
41	1066	56.2	1675	10	AF042059	AF042059 Rattus no
42	1056.4	55.7	1512	10	RNAJ2603	AJ002603 Rattus no
43	1053.2	55.5	1512	10	AB012721	AB012721 Rattus no
44	1043.6	55.0	1512	10	AY054413	AY054413 Mus muscu
45	932	49.1	1215	9	AF074599	AF074599 Homo sapi
46	908.4	47.9	1504	10	AF042060	AF042060 Rattus no
47	879.4	46.3	3163	5	AF113513	AF113513 Sturnus v
48	844.4	44.5	1558	10	AF042061	AF042061 Rattus no
49	823.4	43.4	3539	5	AB036415	AB036415 Gallus ga
50	811.2	42.7	840	9	AF119229	AF119229 Macaca mu
51	764.8	40.3	1419	5	AF045149	AF045149 Coturnix
52	662.4	34.9	1795	5	AF147746	AF147746 Squalus a
53	616.6	32.5	751	4	CFA313196	AJ313196 Canis fam
54	519.2	27.4	3061	5	AB003356	AB003356 Anguilla
55	500	26.3	1937	5	DRE414567	AJ414567 Danio rer
56	485.2	25.6	2018	5	AF349414	AF349414 Danio rer
57	484.2	25.5	2283	5	AF061269	AF061269 Carassius
58	479.4	25.3	1909	5	AB083064	AB083064 Cyprinus
59	469.6	24.7	2508	5	OMY289883	AJ289883 Oncorhync
60	468.4	24.7	2850	5	AB070630	AB070630 Paralicht
61	460.2	24.2	2276	5	AB070901	AB070901 Oryzias l
62	450.6	23.7	2180	5	AF136980	AF136980 Sparus au
63	449	23.7	4306	5	TNU75605	U75605 Tilapia nil
64	441.2	23.2	1772	5	AF298182	AF298182 Micropogo
c 65	438	23.1	127709	2	CNS05701	AL352983 Homo sapi

c	66	438	23.1	143340	9	AF215937	AF215937 Homo sapi
	67	438	23.1	176257	9	CN501RRD	AL161756 Human chr
	68	438	23.1	325791	6	AX234657	AX234657 Sequence
	69	432.2	22.8	3804	5	AF185568	AF185568 Ictalurus
	70	431.6	22.7	2650	5	AF177465	AF177465 Carassius
	71	427	22.5	2362	5	AF298181	AF298181 Micropogo
	72	426.2	22.5	716	4	AF257109	AF257109 Ovis arie
	73	417.4	22.0	1925	5	DRE414566	AJ414566 Danio rer
	74	395.8	20.9	2764	6	AX376687	AX376687 Sequence
	75	395.8	20.9	2770	5	ORZMER	D28954 Oryzias sp.
	76	395.4	20.8	2390	5	AF298183	AF298183 Micropogo
	77	395	20.8	626	10	AB076607	AB076607 Mesocrice
	78	392.8	20.7	2792	5	ZP1ER	L79911 Taeniopygia
	79	390.8	20.6	1474	6	AX066402	AX066402 Sequence
	80	390.8	20.6	1788	6	AX411751	AX411751 Sequence
	81	390.8	20.6	4963	6	AR029418	AR029418 Sequence
	82	390.8	20.6	4963	6	115368	115368 Sequence 16
	83	390.8	20.6	4963	6	156760	156760 Sequence 9
	84	389.4	20.5	1728	6	E39777	E39777 Estrogen re
	85	389.4	20.5	1863	6	E39778	E39778 Estrogen re
	86	389.2	20.5	2092	6	AX474709	AX474709 Sequence
	87	389.2	20.5	2092	6	108538	108538 Sequence 1
	88	389.2	20.5	2092	9	HUMERMCF	M12674 Human estro
	89	389.2	20.5	6450	6	AR153585	AR153585 Sequence
	90	389.2	20.5	6450	9	HSERR	X03635 Human mRNA
	91	389.2	20.5	6610	12	AF061181	AF061181 Mammalian
	92	388.2	20.5	2085	5	GGERR	X03805 Chicken mRN
	93	388	20.4	1788	4	SSESTREC	Z37167 S.scrofa mR
	94	385	20.3	2552	5	DRE275911	AJ275911 Danio rer
	95	384.4	20.2	5078	5	TNU75604	U75604 Tilapia nil
	96	384.2	20.2	5653	4	AY033393	AY033393 ovis arie
	97	382.6	20.2	2423	5	AF326201	AF326201 Halicheer
	98	382.4	20.1	3416	5	AF136979	AF136979 Sparus au
	99	382	20.1	1966	4	OAESTREC	Z49257 O.aries mRN
	100	381.2	20.1	2821	5	AB037185	AB037185 Danio rer
	101	380.8	20.1	2369	5	SAU6039	AJ006039 Sparus au
	102	379.2	20.0	3500	5	AB007453	AB007453 Chrysophr
	103	378	19.9	3439	5	AF349412	AF349412 Danio rer
	104	375.8	19.8	1880	10	RNESTROR	X61098 Rat mRNA fo
	105	375.8	19.8	2090	10	S52128	S52128 estrogen re
	106	375.8	19.8	2090	10	RNERR	Y00102 Rat mRNA fo
	107	375.2	19.8	1223	6	AX066401	AX066401 Sequence
	108	375.2	19.8	3746	4	AF124093	AF124093 Equus cab
	109	374.8	19.7	1764	5	AB055220	AB055220 Caiman cr
	110	366.2	19.3	1978	10	AF181077	AF181077 Mesocrice
	111	365	19.2	2796	5	AB070629	AB070629 Paralicht
	112	363	19.1	2078	10	MUSESTRU	M38651 Mouse estro
	113	359.8	19.0	2102	5	CGESTREC	X84743 C.gariepinu
	114	354.4	18.6	2087	5	AY055725	AY055725 Carassius
	115	354.4	18.6	6041	5	AF253505	AF253505 Ictalurus
	116	354.4	18.6	6756	5	AF061275	AF061275 Ictalurus
c	117	354.4	18.6	8762	5	AF253507	AF253507 Ictalurus
	118	351.8	18.5	1746	5	AB055221	AB055221 Chemidoph
	119	342.2	18.0	1955	5	OMY242741	AJ242741 Oncorhync
	120	342.2	18.0	2049	5	OMY242740	AJ242740 Oncorhync

ALIGNMENTS

```
RESULT 1
A65127
LOCUS       A65127                  1898 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 24 from Patent EP0798378.
ACCESSION   A65127
VERSION     A65127.1  GI:4540991
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1898)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378 A 24 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     Other publication: CA 2200423 19970926
            Other publication: AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       452 a    506 c    501 g    439 t
ORIGIN

  Query Match          100.0%;  Score 1898;  DB 6;  Length 1898;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1898;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGACTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCACATCTGTATGCGGAACCTCAAAAGACTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660

Qy        661 AAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780

Qy        781 CGACGAGCAGCTGCACTGTGCCCGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGACGAGCAGCTGCACTGTGCCCGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 GCGGGAGGTGCTGCTGGAGGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACGGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACGGAGGCCTCCATGAT 960

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCCACCAAGTGCGGCTCTTCGAGAGCTGTTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCCACCAAGTGCGGCTCTTCGAGAGCTGTTC 1080

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAGCTCAT 1140

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGCGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTTTGCTCCAGATCTTGTTCTGGACAGCGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680

Qy       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGCACCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGGAGCC 1740

Qy       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800

Qy       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTTCCCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTTCCCT 1860

Qy       1861 TGTTAGCAGAGGGACATTTGAATCGAGCGTTTCCACAC 1898
              |||||||||||||||||||||||||||||||||||||||
Db       1861 TGTTAGCAGAGGGACATTTGAATCGAGCGTTTCCACAC 1898


RESULT 2
AX323078
```

```
LOCUS       AX323078                1898 bp    DNA     linear   PAT 07-JAN 2002
DEFINITION  Sequence 24 from Patent EP1162264.
ACCESSION   AX323078
VERSION     AX323078.1  GI:18093960
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Messelman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 24 12 DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      452 a    506 c    501 g    439 t
ORIGIN

  Query Match           100.0%;  Score 1898;  DB 6;  Length 1898;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1898;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
```

```
Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC  780

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT  840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT  840

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC  900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC  900

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT  960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT  960

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260

Qy       1261 ACAATATCTCTGTGTCAAGCCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGAATATCTCTGTGTCAAGCCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTCACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTCACCGA 1380

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGGCCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGGCCCT 1440

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680

Qy       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGGAGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGGAGCC 1740

Qy       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800

Qy       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTTCCCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTTCCCT 1860

Qy       1861 TGTTAGCAGAGGGACATTTGAATCGACGTTTCCACAC 1898
              |||||||||||||||||||||||||||||||||||||
Db       1861 TGTTAGCAGAGGGACATTTGAATCGACGTTTCCACAC 1898
```

```
RESULT 3
AB006590
LOCUS       AB006590                1740 bp    mRNA    linear   PRI 05-FEB 1999
DEFINITION  Homo sapiens mRNA for estrogen receptor beta, complete cds.
ACCESSION   AB006590
VERSION     AB006590.1  GI:2911151
KEYWORDS    estrogen receptor beta.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogawa,S., Inoue,S., Watanabe,T., Hiroi,H., Orimo,A., Hosoi,T.,
            Ouchi,Y. and Muramatsu,M.
  TITLE     The complete primary structure of human estrogen receptor beta (hER
            beta) and its heterodimerization with ER alpha in vivo and in vitro
  JOURNAL   Biochem. Biophys. Res. Commun. 243 (1), 122-126 (1998)
  MEDLINE   98139878
REFERENCE   2  (bases 1 to 1740)
  AUTHORS   Ogawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1997) Sumito Ogawa, Saitama Medical School,
            Department of 2nd Biochemistry; 38 Morohongo, Moroyama, Iruma-gun,
            Saitama 350-0495, Japan (E-mail:suogawa@saitama-med.ac.jp,
            Tel:81-492-76-1490, Fax:81-492-94-9751)
FEATURES             Location/Qualifiers
     source          1. .1740
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
     CDS             99. .1691
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="BAA24953.1"
                     /db_xref="GI:2911152"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVIWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEARSLEHTLPVNRETLKRKVSGNRGASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRKERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT
                     GSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT      425 a    463 c    459 g    393 t
ORIGIN

  Query Match             90.5%;  Score 1718;  DB 9;  Length 1740;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1718;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGACAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 82

Qy         61 GCTGTTATCTCAACACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 GCTGTTATCTCAACACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 142

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 202

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        203 CTATGTAGACAGCCACCATGAAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 262

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 322

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 382

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 442

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 502

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTCCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTCCAGCGATTA 562

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 622

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 682

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAACTGGGAATGCT 742

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 802

Qy        781 CGACCACCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCGCCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCGCCCCCGAGT 862

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 922

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 982

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATCATCAGCTGGCCCAACAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATCATCAGCTGGCCCAAGAA 1042

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTGTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTGTTGGAGAGCTGTTG 1102

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1162

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAACGAATTCTGCA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAACGAATTCTGCA 1222

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1223 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1282

Qy       1261 ACAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1283 ACAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1342

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1343 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1402

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1403 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1462

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1463 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1522

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1523 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1582

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1583 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1642

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1643 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1702

Qy       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1718
              ||||||||||||||||||||||||||||||||||||||
Db       1703 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1740


RESULT 4
AX234658
LOCUS       AX234658                2011 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 2 from Patent WO0162793.
ACCESSION   AX234658
VERSION     AX234658.1  GI:15593629
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Kalush,F., Cassel,M.J., Hwang,S.S. and Winn-Deen,E.S.
  TITLE     Estrogen receptor beta variants and methods of detection thereof
  JOURNAL   Patent: WO 0162793-A 2 30 AUG 2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1..2011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      470 a    541 c    561 g    439 t
ORIGIN

  Query Match             87.9%;  Score 1669;  DB 6;  Length 2011;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1669;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 402

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 462

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 522

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 582
```

```
Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 642

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 702

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 762

Qy        421 CACCTTACCTGTAAACAGAGACACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CACCTTACCTGTAAACAGAGACACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 822

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              |||||||||||||||||||||||||||||||.|||||||||||||||||||||||||||
Db        823 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 882

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 942

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 1002

Qy        661 AAACCGGCGCAAGAGCTGCCAGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 AAACCGGCGCAAGAGCTGCCAGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 1062

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 1122

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 1182

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 1242

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              |||||||||||||||.|||||||||||||||||||||||||||||||||||||||||||
Db       1243 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 1302

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1303 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1362

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTC 1422

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1482

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1542

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1602

Qy       1261 ACAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 ACAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1662

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1722

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1782

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1842

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1902

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1962

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 2011
```

```
RESULT 5
AF051427
LOCUS       AF051427                2011 bp    mRNA    linear   PRI 14-JUL-1999
DEFINITION  Homo sapiens estrogen receptor beta mRNA, complete cds.
ACCESSION   AF051427
VERSION     AF051427.1  GI:2970563
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Moore,J.T., McKee,D.D., Slentz-Kesler,K., Moore,L.B., Jones,S.A.,
            Horne,E.L., Su,J.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Cloning and characterization of human estrogen receptor beta
            isoforms
  JOURNAL   Biochem. Biophys. Res. Commun. 247 (1), 75-78 (1998)
  MEDLINE   98300286
   PUBMED   9636657
REFERENCE   2  (bases 1 to 2011)
  AUTHORS   Moore,J.T., McKee,D.D., Moore,L.B., Jones,S.A., Su,J.-L.,
            Slentz-Kesler,K., Horne,E.L., Kliewer,S.A., Lehmann,J.M. and
            Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..2011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             419..2011
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAC05985.1"
                     /db_xref="GI:2970564"
                     /translation="MDIKNSPSSLNSPSSYNCSQSIIPEFHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SRRERCGYRLVRKQKSADEQLRCACKAKREGGHAPRVREELILDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT
                     GSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT      470 a     541 c     561 g     439 t
ORIGIN
```

```
  Query Match             87.9%;  Score 1669;  DB 9;  Length 2011;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1669;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 402

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 462

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 522

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 582

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 642

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CCCAAATGTCTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 702

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 762

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCCCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCCCCAG 822

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 882

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 942

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 1002

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 1062

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 1122

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 1182

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 1242

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1243 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 1302

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1303 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1362

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1422

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1482

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1542

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1602

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1662

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1722

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1782

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1842

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1902

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1962

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 2011


RESULT 6
AX029400
LOCUS       AX029400               1647 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9912961.
ACCESSION   AX029400
VERSION     AX029400.1  GI:10190195
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1647)
  AUTHORS   Wilkinson,H.
  TITLE     Estrogen receptor
  JOURNAL   Patent: WO 9912961-A 2 18-MAR-1999;
            MERCK & CO INC (US) ; WILKINSON HILARY (US)
FEATURES             Location/Qualifiers
     source          1. .1647
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      401 a    438 c    436 g    372 t
ORIGIN

  Query Match             86.2%;  Score 1635.6;  DB 6;  Length 1647;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1638;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAA 87
```

```
                 [ 31    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 CTTTGTAGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAA 65

Qy         88 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 125

Qy        148 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 185

Qy        208 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 245

Qy        268 CTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGG 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 CTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGG 305

Qy        328 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 365

Qy        388 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 425

Qy        448 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 485

Qy        508 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 545

Qy        568 GTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATAT 627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 GTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATAT 605

Qy        628 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTG 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTG 665

Qy        688 CCCACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGACAGATG 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 CCCACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGACAGATG 725

Qy        748 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 785

Qy        808 GGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 GGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAC 845

Qy        868 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 905

Qy        928 CCCCAGTGCGCCCTTCACGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 CCCCAGTGCGCCCTTCACGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 965

Qy        988 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1025

Qy       1048 GTTCGACCAAGTGCGGCTCTTGGAGACGTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 GTTCGACCAAGTGCGGCTCTTGGAGACGTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1085

Qy     ' 1108 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1086 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1145

Qy       1168 GGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 GGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1205

Qy       1228 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGAT 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGAT 1265

Qy       1288 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1325

Qy       1348 GAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 GAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1385

Qy       1408 CGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 CGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1445

Qy       1468 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1505

Qy       1528 GGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAA 1587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 GGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAA 1565

Qy       1588 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTC 1647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTC 1625

Qy       1648 CCAGAACCCACAGTCTCAGTGA 1669
              ||||||||||||||||||||||
Db       1626 CCAGAACCCACAGTCTCAGTGA 1647
```

```
RESULT 7
HSRNAERB
LOCUS       HSRNAERB                1560 bp    mRNA    linear   PRI 05-SEP-1996
DEFINITION  H.sapiens mRNA for estrogen receptor.
ACCESSION   X99101
VERSION     X99101.1  GI:1518262
KEYWORDS    estrogen receptor.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1560)
  AUTHORS   Mosselman,S., Polman,J. and Dijkema,R.
  TITLE     ER beta: identification and characterization of a novel human
            estrogen receptor
  JOURNAL   FEBS Lett. 392 (1), 49-53 (1996)
  MEDLINE   96354875
   PUBMED   8769313
REFERENCE   2  (bases 1 to 1560)
  AUTHORS   Mosselman,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-1996) S. Mosselman, N.V.Organon, Biotechnology
            and Biochemistry, PO box 20 oss, Molenstraat 110 oss, 5340 BH,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .1560
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /tissue_type "testis"
                     /clone_lib "Clontech: HL1010B"
                     /clone_lib "Clontech: 7414 1"
                     /dev_stage="adult"
     CDS             19. .1452
                     /codon_start 1
```

```
                     /product="estrogen receptor beta"
                     /protein_id="CAA67555.1"
                     /db_xref="GI:1518263"
                     /db_xref="SPTREMBL:Q92731"
                     /translation="MNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSH
                     LYAEPQKSPWCEARSLETTLPVNRFTLKRKVSGNPCASPVTGPGSKEDAHFCAVCSDY
                     ASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVG
                     MVKCGSRRERCGYRLVRRQRSADFQLHCAGKAKRSGCHAPVRELILDALSPFQLVIT
                     LLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVR
                     LLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRF
                     RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSG
                     ISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGC
                     KSSITGSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT      367 a    408 c    445 g    340 t
ORIGIN

  Query Match             82.1%;  Score 1558;  DB 9;  Length 1560;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1558;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        220 CTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTGCAAGGTGG 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 CTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGG 62

Qy        280 GCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCC 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 GCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCC 122

Qy        340 TTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTG 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTG 182

Qy        400 TGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAAACGAACGT 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGT 242

Qy        460 TAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAACAGGGATGCTCACTT 519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTT 302

Qy        520 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGG 579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGG 362

Qy        580 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 639
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 422

Qy        640 AAATCAGTGTACAATCGATAAAAACCGGGCCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 AAATCAGTGTACAATCGATAAAAACCGGGCCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 482

Qy        700 GTGTTACGAAGTGGGAATGGTCAAGTGTGGCTCCCCGAGAGAGAGATGTGGGTACCGCCT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GTGTTACGAAGTGGGAATGGTCAAGTGTGGCTCCCCGAGAGAGAGATGTGGGTACCGCCT 542

Qy        760 TGTGCGGAGACAGAGAAGTGCCGAGCAGCAGCTGCACTGTGCCCGCAAGGCCAAGACAAG 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TGTGCGGAGACAGAGAAGTGCCGAGCAGCAGCTGCACTGTGCCCGCAAGGCCAAGACAAG 602

Qy        820 TGGCGGCCACGGCGCCCCGAGTGCGGGGAGCTGCTGCTGGAGCGCCTGAGCGCCGACCAGCT 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 TGGCGGCCACGGCGCCCCGAGTGCGGGGAGCTGCTGCTGGAGCGCCTGAGCGCCGACCAGCT 662

Qy        880 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCC 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCC 722

Qy        940 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGCTGGTCCAC 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGCTGGTCCAC 782

Qy       1000 CATGATCAGCTGGGCCAAGAAGATCCCGGGCTTTGTGGACCTCACCCTGCATGACCAAGT 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 CATGATCAGCTGGGCCAAGAAGATTCCGGGCTTTGTGGACCTCACCCTGCATGACCAAGT 842

Qy       1060 GCACCTCTTGCAGAGCTGTTGGATGGAGGTGTTAATGATGGGCTGATCGTGGCGCTCAAT 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 GCACCTCTTGCAGAGCTGTTGGATGGAGGTGTTAATGATGGGCTGATCGTGGCGCTCAAT 902

Qy       1120 TGACCACCCCGGCCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAGGGAA 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 TGACCACCCCGGCCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAGGGAA 962

Qy       1180 ATGCGTAGAAGCAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCG 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 ATGCGTAGAAGCAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCG 1022

Qy       1240 AGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTC 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1023 AGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTC 1082

Qy       1300 CAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCA 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 CAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCA 1142

Qy       1360 CTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTC 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1143 CTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTC 1202

Qy       1420 CCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGC 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 CCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGC 1262

Qy       1480 GAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGGTGCAAAAAATGTGCTCCCAGTGTA 1539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 GAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGGTGCAAAAAATGTGCTCCCAGTGTA 1322

Qy       1540 TGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCAC 1599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 TGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCAC 1382

Qy       1600 GGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACA 1659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 GGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACA 1442

Qy       1660 GTCTCAGTGACGCCTGGCCCTGAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTG 1719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 GTCTCAGTGACGCCTGGCCCTGAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTG 1502

Qy       1720 AACTCCAGTGTGTCAGGAGCCTGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTGG 1777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1503 AACTCCAGTGTGTCAGGAGCCTGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTGG 1560


RESULT 8
CJERB
LOCUS       CJERB                   1651 bp    mRNA    linear   PRI 11-NOV-1999
DEFINITION  C.jacchus mRNA for estrogen receptor beta.
ACCESSION   Y09372
VERSION     Y09372.2  GI:6433823
KEYWORDS    DNA binding; estrogen receptor; steroid hormone receptor;
            transactivation.
SOURCE      Callithrix jacchus.
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
            Callithrix.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Gaughan,J.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1651)
  AUTHORS   Gaughan,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1996) J. Gaughan, Medical Research Council,
            Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh, EH3
```

```
            9EW, UK
REFERENCE   4  (bases 1 to 1651)
  AUTHORS   Scobie,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) G. Scobie., Medical Research Council,
            Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh, EH3
            9EW, UK
COMMENT     On Nov 15, 1999 this sequence version replaced gi:1669661.
            Related entries: X99101 & U57439.
FEATURES             Location/Qualifiers
     source          1. .1651
                     /organism "Callithrix jacchus"
                     /db_xref "taxon:9483"
                     /clone "pGEM T vector"
                     /tissue_type "prostate"
                     /tissue_type "ovary"
                     /tissue_type "testis"
     CDS             52. .1644
                     /function "steroid hormone receptor"
                     /codon_start=1
                     /product "estrogen receptor beta"
                     /protein_id "CAA70546.2"
                     /db_xref="GI:6433824"
                     /db_xref="SPTREMBL:Q95171"
                     /translation "MDIKNSPSSESNSPSSYNFGQSIPLPFHGPIYIPSSYVESHHEYP
                     AMTFYSPAVMNYSIPSSVINLEEGPGRQITSPNMLWSTPGHLSPLAVHHQLSHLYAEP
                     QKSPWCEARSLEHTLPVSRETLKRKVSGNHCASPVTGPSSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRRERCGYRLVRRQGNAEEQLHCAGKAKRSGGHVPRVKELLLSALSPEQLVLTLLEAE
                     PPHVLISRPSVPFTEASMMMSLTKLADEELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WLEVLMVGLMWRSIDHPGKLIFAPNLLLDRDEGKCVEGILEVFDMLLATTSRFRELKL
                     QHKEYLCVKAMVLLNSQYDRLVTATQDAESSQKLAHLLNAVTDALVWVIAKSGFSSQQ
                     QSVRLANLLMLLSHIRHASNKGMEHLLSMKCKNVVPVYDLLLEMMNAHVVRGCKSSIT
                     GSECSPAEDSKSTECSQNPQSP"
BASE COUNT      497 a    434 c    443 g    377 t
ORIGIN

  Query Match             78.7%;  Score 1494.2;  DB 9;  Length 1651;
  Best Local Similarity   94.1%;  Pred. No. 0;
  Matches 1558;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy         26 GACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATA 85
              | ||||||||||||||| || ||||||| ||  ||||||||||||||||||||||||||
Db          1 GCCCTTTGTGCCTCTTCTCGCGAGGTGTTATCCTAGCTGTTATCTCAAGACATGGATATA 60

Qy         86 AAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTA 145
              |||||||||||||| ||||||||||||||||| |||||||| |   ||||||||||||||
Db         61 AAAAACTCACCATCCAGCCTTAATTCTCCTTCTTCCTACAATTTTGGTCAATCCATCTTA 120

Qy        146 CCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATAT 205
              ||| |||||||||| | ||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCTTGGAGCACGGCCCTATATACATACCTTCCTCCTATGTAGAGAGCCACCATGAATAT 180

Qy        206 CCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACT 265
              || ||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db        181 CCGGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAGTGTTACT 240

Qy        266 AACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCT 325
              ||||||||||  |||||||||||||||| |||||||||||| || |||| |||||||
Db        241 AACTTGGAAGAGGGGCCTGGTCGGCAGATCACAAGCCCAAATATGCTGTGGTCAACACCT 300

Qy        326 GGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAA 385
              || ||||||||||||||  ||||||| |||||||||||||||||| |||||||||||||
Db        301 GGCCACCTTTCTCCTTTAGCTGTCCATCACCAGTTATCACATCTGTATGCAGAACCTCAA 360

Qy        386 AAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACA 445
              |||||||||||||||||||||||||| |||||||| |||||||||||| ||||||||||
Db        361 AAGAGTCCCTGGTGTGAAGCAAGATCACTAGAACATACCTTACCTGTAAGCAGAGAGACA 420

Qy        446 CTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAG 505
              |||||||||| |||||||||||| ||| ||||||||||||||||||| ||||||||
Db    '   421 CTGAAAAGGAAAGTTAGTGGGAACCATTGTGCCAGCCCTGTTACTGGTCCAAGTTCAAAG 480

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        481 AGGGATGCTCACTTCTGTGCAGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 540

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 600

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 660

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||||||||||||||| ||||||||||||||| ||||||||||||||||||||||||||
Db        661 TGCCGACTTCGGAAGTGCTACGAAGTGGGAATGGTAAAGTGTGGCTCCCGGAGAGAGAGA 720

Qy        746 TGTGGGTACCGGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGC 805
              |||||||||||||||||||||||||||| ||| |||||| |||||||||| || ||||||
Db        721 TGTGGGTACCGGCCTTGTGCGGAGACAGGGAAATGCCGAGGAGCAGCTGCATTGCGCCGGC 780

Qy        806 AAGGCCAACAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTG 865
              |||||||||||||||| || |||| |||||||||||||||||||||||  |||||||
Db        781 AAGGCCAAGAGAAGTGGTGGTCACGTGCCCCGAGTGCGGGAGCTGCTGCTGAGCGCCCTG 840

Qy        866 AGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCCGAGCCGCCCCCATGTGCTGATCAGC 925
              |||||||||||||||||||||||||||||||||||| || ||||||||||||||||||||
Db        841 AGCCCCGAGCAGCTGGTGCTCACCCTCCTGGAGGCCGAACCGCCCCCATGTGCTGATCAGC 900

Qy        926 CGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGAC 985
              |||||||| | ||||||||| |||||||||||||||||||||||||||| |||||||||
Db        901 CGCCCCAGCCTGCCCTTCACTGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCCGAC 960

Qy        986 AAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGC 1045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGC 1020

Qy       1046 CTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTG 1105
              ||||| |||||||||||||||||||||||| ||| ||||||||||||| ||||||||
Db       1021 CTGTTGGACCAAGTGCGGCTCTTGGAGAGCTGCTGGTTGGAGGTGTTAATGGTGGGGCTG 1080

Qy       1106 ATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGAC 1165
              |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||
Db       1081 ATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAAATCTCATTCTGGAC 1140

Qy       1166 AGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACT 1225
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAGTCTTTGACATGCTCCTGGCAACT 1200

Qy       1226 ACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATG 1285
              |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db       1201 ACTTCAAGGTTTCGAGAGTTGAAACTCCAACACAAAGAGTATCTCTGTGTCAAGGCCATG 1260

Qy       1286 ATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGC 1345
              ||||||||||||||      |||||||||||||||| ||||||||||||| ||||||
Db       1261 GTCCTGCTCAATTCCCAGTATGACCCTCTGGTCACAGCAACCCAGGATGCTGAGAGCAGC 1320

Qy       1346 CGGAAGCTGGCTCACTTGCTGAACGCCCGTGACCGATGCTTTGGTTTGGGTGATTGCTAAG 1405
              | |||||||| |||||||||| |||||||| || |||||||| |||||||| || |||
Db       1321 CAGAAGCTGGCCCACTTGCTGAATGCCGTGACTGACGCTTTGGTCTGGGTGATCGCTAAG 1380

Qy       1406 AGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCC 1465
              || ||| |||||||||||||||| ||  ||||||||||||||||||||||||||||||
Db       1381 AGTGGCTTCTCCTCCCAGCAGCAGTCGGTGCGCCTGGCTAACCTCCTGATGCTCCTGTCC 1440

Qy       1466 CACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAAT 1525
              ||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       1441 CACATCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAAT 1500

Qy       1526 GTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGC 1585
              |||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db       1501 GTGGTCCCAGTGTATGACCTGCTGCTGGAGATGATGAATGCCCACGTGGTTCGCGGGTGC 1560

Qy       1586 AAGTCCTCCATCAGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGC 1645
```

```
                                                
```

```
Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1664 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1723

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1724 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1783

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1784 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1824


RESULT 10
AF060555
LOCUS       AF060555                2745 bp    mRNA    linear   PRI 29-APR-1998
DEFINITION  Homo sapiens estrogen receptor beta 3 isoform mRNA, complete cds.
ACCESSION   AF060555
VERSION     AF060555.1  GI:3091285
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2745)
  AUTHORS   Moore,J.T., McKee,D.D., Slentz-Kesler,K., Moore,L.B., Jones,S.A.,
            Su,J.L., Horne,E.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..2745
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             471..2012
                     /note="splice variant"
                     /codon_start=1
                     /product="estrogen receptor beta 3 isoform"
                     /protein_id="AAC15234.1"
                     /db_xref="GI:3091286"
                     /translation="MDIKNSPSSINSPSSYNCSQSTLPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEAASIFHTLPVNRETLKRKVSGNPCASPVTGPGSKEDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASSLSLSWRLFMLREASCHGVRQTPGGAHMSVSRSRSFEAC
                     QQPRE"
BASE COUNT      741 a    665 c    690 g    649 t
ORIGIN

  Query Match          78.1%;  Score 1482;  DB 9;  Length 2745;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 454

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 514

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 574

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 634

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGCAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 TTACAGCATTCCCAGCAATGTCACTAACTTGCAAGGTGGGCCTGGTCGGCAGACCACAAG 694

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 754

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 814

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        815 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 874

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        875 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 934

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        935 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 994

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        995 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 1054

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1055 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 1114

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1115 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 1174

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1175 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 1234

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1235 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 1294

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1295 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 1354

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1355 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1414

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1415 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1474

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1475 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1534

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1535 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1594

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1595 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1654

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1655 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1714

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
```

```
Db       1715 ACCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAAGGCCCTCACCCA 1774

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATCGCCCT 1440

Db       1775 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATCGCCCT 1834

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCCAG 1482

Db       1835 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCCAG 1876


RESULT 11
AB006589
LOCUS       AB006589                3593 bp    mRNA    linear   PRI 28-JUL-1998
DEFINITION  Homo sapiens mRNA for estrogen receptor beta cx, complete cds.
ACCESSION   AB006589
VERSION     AB006589.1  GI:3345671
KEYWORDS    estrogen receptor beta cx.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogawa,S., Inoue,S., Watanabe,T., Orimo,A., Hosoi,T., Ouchi,Y. and
            Muramatsu,M.
  TITLE     Molecular cloning and characterization of human estrogen receptor
            betacx: a potential inhibitor ofestrogen action in human
  JOURNAL   Nucleic Acids Res. 26 (15), 3505-3512 (1998)
  MEDLINE   98337908
REFERENCE   2  (bases 1 to 3593)
  AUTHORS   Ogawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1997) Sumito Ogawa, Saitama Medical School,
            Department of 2nd Biochemistry; 38 Morohongo, Moroyama, Iruma-gun,
            Saitama 350-0495, Japan (E-mail:sumogawa@saitama-med.ac.jp,
            Tel:81-492-76-1490, Fax:81-492-94 9751)
FEATURES             Location/Qualifiers
     source          1..3593
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             1276..2763
                     /codon_start=1
                     /product="estrogen receptor beta cx"
                     /protein_id="BAA31966.1"
                     /db_xref="GI:3345672"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPDRQTTSPNVIWPIPCHLSPIVVHPQLSHIYAEP
                     QKSPWCFARSLPHTLPVNPETIKPKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASEKASQTLTSFGMKMETLLPEATMEQ"
BASE COUNT      971 a    811 c    957 g    854 t
ORIGIN

  Query Match             78.0%;  Score 1480.4;  DB 9;  Length 3593;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60

Db       1200 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 1259

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120

Db       1260 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 1319

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180

Db       1320 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 1379

Qy         181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240

Db       1380 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 1439

Qy         241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300

Db       1440 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 1499

Qy         301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360

Db       1500 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 1559

Qy         361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420

Db       1560 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 1619

Qy         421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480

Db       1620 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 1679

Qy         481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540

Db       1680 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 1739

Qy         541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600

Db       1740 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 1799

Qy         601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660

Db       1800 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 1859

Qy         661 AAACGGGCCCAACAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTACCAACTGGCAATGCT 720

Db       1860 AAACGGGCCCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACCAACTGGCAATGGT 1919

Qy         721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780

Db       1920 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 1979

Qy         781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840

Db       1980 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 2039

Qy         841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900

Db       2040 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 2099

Qy         901 TGAGCCGCCCCACATGTCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960

Db       2100 TGAGCCGCCCCACATGTCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 2159

Qy         961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020

Db       2160 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 2219

Qy        1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080

Db       2220 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTACGGCTCTTGGAGAGCTGTTG 2279

Qy        1081 GATGGAGCTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140

Db       2280 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 2339

Qy        1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200

Db       2340 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 2399

Qy        1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260

Db       2400 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 2459
```

```
Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2460 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 2519

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2520 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 2579

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2580 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 2639

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       2640 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 2681


RESULT 12
A61586
LOCUS       A61586                  1460 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 4 from Patent WO9709348.
ACCESSION   A61586
VERSION     A61586.1  GI:4715917
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1460)
  AUTHORS   Kuiper,G.G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348-A 4 13 MAR-1997;
            KAROBIO AB (SE)
FEATURES             Location/Qualifiers
     source          1..1460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match           76.6%;  Score 1453.6;  DB 6;  Length 1460;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1456;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        210 CCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 269
              | ||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTATGACATTCTACAGTCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 60

Qy        270 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 329
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db         61 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 120

Qy        330 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 180

Qy        390 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 240

Qy        450 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 300

Qy        510 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGCATATCACTATGGAGTCTGGT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGCATATCACTATGGAGTCTGGT 360

Qy        570 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 420

Qy        630 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 689
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 480

Qy        690 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTG 749
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTG 540

Qy        750 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 809
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 600

Qy        810 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 869
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 660

Qy        870 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCCATGTGCTGATCAGCCGCC 929
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCCATGTGCTGATCAGCCGCC 720

Qy        930 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 989
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 780

Qy        990 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 1049
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 840

Qy       1050 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 1109
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 900

Qy       1110 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 1169
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 960

Qy       1170 ATGAGGCGAAATGCGTAGAAGGAATTCTCGAAATCTTTCACATGCTCCTGCCAACTACTT 1229
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTT 1020

Qy       1230 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1289
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1080

Qy       1290 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1349
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1140

Qy       1350 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1409
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1200

Qy       1410 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1469
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1260

Qy       1470 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1529
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1320

Qy       1530 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1589
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1380

Qy       1590 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1649
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1440

Qy       1650 AGAACCCACAGTCTCAGTGA 1669
              |||||| ||||||||||||||
Db       1441 AGAACCTACAGTCTCAGTGA 1460


RESULT 13
```

```
AR075921
LOCUS       AR075921               1460 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 4 from patent US 5958710.
ACCESSION   AR075921
VERSION     AR075921.1  GI:13002667
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1460)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J.-A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710 A 4 28 SEP-1999,
FEATURES             Location/Qualifiers
     source          1. .1460
                     /organism="unknown"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match             76.6%;  Score 1453.6;  DB 6;  Length 1460;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1456;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        210 CCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 269
              | ||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db          1 CTATGACATTCTACAGTCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 60

Qy        270 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 120

Qy        330 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 180

Qy        390 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 240

Qy        450 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 300

Qy        510 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 360

Qy        570 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 420

Qy        630 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGCGCGCAAGAGCTGCCAGGCCTGCC 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGCGCGCAAGAGCTGCCAGGCCTGCC 480

Qy        690 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGACACATCTG 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTG 540

Qy        750 GGTACCGCCTTGTGCGGACACAGACAAGTGCCGACGAGCAGGTGCACTGTGCCGGCAAGG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTACCGCCTTGTGCGGAGACAGACAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 600

Qy        810 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 660

Qy        870 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 720

Qy        930 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 780

Qy        990 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 840

Qy       1050 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGCAGGTGTTAATGATGGGGCTGATGT 1109
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGCAGGTGTTAATGATGGGGCTGATGT 900

Qy       1110 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 1169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 960

Qy       1170 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCGTGGCAACTACTT 1229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCGTGGCAACTACTT 1020

Qy       1230 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1080

Qy       1290 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1140

Qy       1350 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1200

Qy       1410 GCATCTCCTCCCAGCAGCAATCCATGCGGCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCATCTCCTCCCAGCAGCAATCCATGCGGCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1260

Qy       1470 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1320

Qy       1530 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1380

Qy       1590 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1440

Qy       1650 AGAACCCACAGTCTCAGTGA 1669
              |||||| ||||||||||||
Db       1441 AGAACCTACAGTCTCAGTGA 1460


RESULT 14
AX022458
LOCUS       AX022458               1460 bp    DNA     linear   PAT 07 SEP 2000
DEFINITION  Sequence 4 from Patent EP0935000.
ACCESSION   AX022458
VERSION     AX022458.1  GI:10046069
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FEATURES             Location/Qualifiers
     source          1. .1460
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match             76.6%;  Score 1453.6;  DB 6;  Length 1460;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1456;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy        210 CCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 269
              1 ||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTATGACATTCTACAGTCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 60

Qy        270 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 120

Qy        330 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 180

Qy        390 GTCCCTGGTGTCAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCCCTGGTGTCAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 240

Qy        450 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 300

Qy        510 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 360

Qy        570 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 420

Qy        630 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 480

Qy        690 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTG 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTG 540

Qy        750 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 600

Qy        810 CCAAGAGAAGTGGCGGCCACGGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAAGAGAAGTGGCGGCCACGGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 660

Qy        870 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 720

Qy        930 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 989
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 780

Qy        990 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 840

Qy       1050 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 1109
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 900

Qy       1110 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 1169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 960

Qy       1170 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTT 1229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTT 1020

Qy       1230 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1080

Qy       1290 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1349
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1140

Qy       1350 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1409
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1200

Qy       1410 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1469
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1260

Qy       1470 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1529
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1320

Qy       1530 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1589
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1380

Qy       1590 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1649
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1440

Qy       1650 AGAACCCACAGTCTCAGTGA 1669
              |||||| ||||||||||||||
Db       1441 AGAACCTACAGTCTCAGTGA 1460


RESULT 15
A65104
LOCUS       A65104                  1434 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 1 from Patent EP0798378.
ACCESSION   A65104
VERSION     A65104.1  GI:4530973
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1434)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 1 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     Other publication CA 2200423 19970926
            Other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1. .1434
                     /organism "unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       347 a    474 c    406 g    308 t
ORIGIN

  Query Match          75.6%;  Score 1434;  DB 6;  Length 1434;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
```

```
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595

Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655

Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715

Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGACACAGAGA 775

Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGACACAGAGA 540

Qy        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835

Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895

Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGCCCCCAGTGCGCCCTTCACCGAGGCCTCC 955

Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGCCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015

Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCGCGCTCTTGGAGAGC 1075

Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCGCGCTCTTGGAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135

Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195

Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255

Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315

Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375

Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435

Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1495

Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260

Qy       1496 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1555

Db       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320

Qy       1556 ATGCTGAATGCCCACCGTCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1615

Db       1321 ATGCTGAATGCCCACCGTCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380

Qy       1616 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669

Db       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434


RESULT 16
AX323055
LOCUS       AX323055                1434 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 1 from Patent EP1162264.
ACCESSION   AX323055
VERSION     AX323055.1  GI:18093942
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 1 12-DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1434
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      347 a    373 c    406 g    308 t
ORIGIN

  Query Match           75.6%;  Score 1434;  DB 6;  Length 1434;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295

Db           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355

Db          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy         356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415

Db         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy         416 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475

Db         181 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy         476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535

Db         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy         536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595

Db         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy         596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655

Db         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy         656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715

Db         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy         716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGACACAGAGA 775

Db         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGACACAGAGA 540

Qy         776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835

Db         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
```

```
Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        896 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260

Qy       1496 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320

Qy       1556 ATGCTGAATGCCCACCGTCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATGCTGAATGCCCACCGTCTTCGCGGGTGCAACTCCTCCATCACGGGGTCCGAGTGCAGC 1380

Qy       1616 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434


RESULT 17
AF393815
LOCUS       AF393815               1640 bp    mRNA    linear   PRI 05-JUL-2001
DEFINITION  Macaca arctoides estrogen receptor beta 2 (ERBeta2) mRNA, complete
            cds.
ACCESSION   AF393815
VERSION     AF393815.1  GI:14600266
KEYWORDS    .
SOURCE      Macaca arctoides.
  ORGANISM  Macaca arctoides
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1640)
  AUTHORS   Scobie,G.A., Wilson,J.A., Millar,M.R., Macpherson,S. and
            Saunders,P.T.
  TITLE     The estrogen receptor beta variant ERBeta cx/ERBeta2 is expressed
            in a wide range of tissues in both Old and New World primates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1640)
  AUTHORS   Scobie,G.A. and Saunders,P.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) Medical Research Council, Human
            Reproductive Sciences Unit, 47 Chalmers Street, Edinburgh EH3 9ET,
            UK
FEATURES             Location/Qualifiers
     source          1..1640
                     /organism="Macaca arctoides"
                     /db_xref="taxon:9540"
     gene            1..1640
                     /gene="ERBeta2"
     CDS             38..1537
                     /gene="ERBeta2"
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta 2"
                     /protein_id="AAK71317.1"
                     /db_xref="GI:14600267"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVESHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTASPNVLWPTPGHLESPLAVHRQLSHLYAEP
                     QKSFWCEARSLEHTLPVNRETLKRKVSGNRCTSPVTSPSSKRDAHFCAVCSDFASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SREFRCGYRLVRRQRSADEQVHCASKAKRSGSHTPLVRELLLDAISPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSNMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHAKAEKASHSLTSEGMEDGDTLLPEAIDGSVTF"
BASE COUNT      406 a    446 c    413 g    384 t      1 others
ORIGIN

  Query Match             72.5%;  Score 1375.8;  DB 9;  Length 1640;
  Best Local Similarity   97.1%;  Pred. No. 0;
  Matches 1401;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy         40 TTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAAAAACTCACCATC 99
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 TTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAAAAATTCACCATC 60

Qy        100 TAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGG 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGG 120

Qy        160 CTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCCAGCCATGACATT 219
              ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        121 CTCTATATACATACCTTCCTCCTATGTAGAGAGCCACCATGAATATCCAGCCATGACATT 180

Qy        220 CTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGG 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGG 240

Qy        280 GCCTGGTCGGCACACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCC 339
              |||||||||||||| ||||||||||||||||||||||||||||||||| ||||||||||
Db        241 GCCTGGTCGGCAGACCGCAAGCCCAAATGTGTTGTGGCCAACACCTGGACACCTTTCTCC 300

Qy        340 TTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTG 399
              |||||  ||||||||||||||||||||| ||||| ||||||||||||||||||||||||
Db        301 TTTAGCAGTCCATCGCCAGTTATCACATCTATATGCAGAACCTCAAAAGAGTCCCTGGTG 360

Qy        400 TGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGT 459
              |||||||||||| ||||||||||||||||||||||||||||||||| ||||||||||||
Db        361 TGAAGCAAGATCACTAGAACACACCTTACCTGTAAACAGAGAGACATTGAAAAGGAAGGT 420

Qy        460 TAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTT 519
              |||||||||||||||| |||||||||||||| ||||| ||||||||||| |||||||||
Db        421 TAGTGGGAACCGTTGCACCAGCCCTGTTACTAGTCCAAGTTCAAAGAGGGACGCTCACTT 480

Qy        520 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGTGAAG 579
              |||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||
```

```
Db        481 CTGCGCTGTCTGCAGTGATTCGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGG 540

Qy        580 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 639
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        541 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGCCATAATGATTATATTTGTCCAGCTAC 600

Qy        640 AAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 699
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        601 AAATCAGTGTACAATCGATAAAAATCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 660

Qy        700 GTGTTACGAAGTGGGAATGGTGAAGTGTGCCTGCCGGAGACAGAGACATGTGGGTACCGGCT 759
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGTTATGAAGTGGGAATGGTGAAGTGTGGCTGCCGGAGACAGAGACATGTGGGTACCGGCT 720

Qy        760 TGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAG 819
              |||||||||||||||||||||||||||||  ||||||| ||| ||||||||||||| ||
Db        721 TGTGCGGCAGACAGACAAGTGCCCACCAGCAAGTGCACTGCGGCAGCAAGGCCAAGAGGAG 780

Qy        820 TGGCGGCCACGGGCGCCGGAGTGCGGGAGCTGCTGTGGACGCCCTGACGCCCGAGCCAGCT 879
              |||| |||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGGCAGCCACAGCGCCCTAGTGCGGGAGCTGCTGTGGACGCCCTGAGCCCGAGCAGCT 840

Qy        880 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGCCCCAGTGCGCC 939
              |||||||||||||||||||||||||||| ||||||| |||||||||||||||| |||||
Db        841 GGTGCTCACCCTCCTGGAGGCTGAGCCACCCCATGTACTGATCAGCGCCCCAGTGCGCC 900

Qy        940 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTCAC 999
              |||||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        901 CTTCACTGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGTTGGTCAC 960

Qy       1000 CATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGT 1059
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGT 1020

Qy       1060 GCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAAT 1119
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1021 GCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGGTGGGGCTGATGTGGCGCTCAAT 1080

Qy       1120 TGACCACCCCGGCAAGCTCATCTTTGCTGGAGTCTTGTTCTGGACAGGATGAGGGCAA 1179
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1081 TGACCACCCCGGCAAGCTCATCTTTGCTGGAGTCTTGTTCTGGATAGGATGAGGGCAA 1140

Qy       1180 ATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCC 1239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCC 1200

Qy       1240 ACACTTAAAACTCCAACACAAACGAATATCTCTGTTCAAGCCATCATCCTGCTCAATTC 1299
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1201 ACAGTTAAAACTCCAACACAAACGAATATCTCTGTTCAAGCCATCATCCTGCTCAATTC 1260

Qy       1300 CAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCCGAAGCTGGCCA 1359
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       1261 CAATATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCCGAAGCTGGCCA 1320

Qy       1360 CTTGCTGAACGCCGTGACCGATGCTTTGGTTGTGGGTGATGCCAAGAGCGCCATCGCCC 1419
              |||||||||||||||||| |||||||| |||||||| |||||||||||||||||||||||
Db       1321 CTTGCTGAACGCCGTGACCGACGCTTTGGTCTGGGTGATGCCAAGAGCGCCATCGCCC 1380

Qy       1420 CCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGC 1479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGC 1440

Qy       1480 GAG 1482
              |||
Db       1441 GAG 1443


RESULT 18
AF393816
LOCUS       AF393816               1566 bp    mRNA    linear   PRI 05-JUL-2001
DEFINITION  Callithrix jacchus estrogen receptor beta 2 (ERBeta2) mRNA,
            complete cds.
ACCESSION   AF393816
VERSION     AF393816.1  GI:14600268
KEYWORDS    .
SOURCE      Callithrix jacchus.
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
            Callithrix.
REFERENCE   1  (bases 1 to 1566)
  AUTHORS   Scobie,G.A., Wilson,J.A., Millar,M.R., Macpherson,S. and
            Saunders,P.T.
  TITLE     The estrogen receptor beta variant ERBeta cx/ERBeta2 is expressed
            in a wide range of tissues in both Old and New World primates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1566)
  AUTHORS   Scobie,G.A. and Saunders,P.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) Medical Research Council, Human
            Reproductive Sciences Unit, 37 Chalmers Street, Edinburgh EH3 9ET,
            UK
FEATURES             Location/Qualifiers
     source          1..1566
                     /organism="Callithrix jacchus"
                     /db_xref="taxon:9483"
     gene            1..1566
                     /gene="ERBeta2"
     CDS             52..1512
                     /gene="ERBeta2"
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta 2"
                     /protein_id="AAK71318.1"
                     /db_xref="GI:14600269"
                     /translation="MDIKNSPSSLNSPSSYNFGQSTFSFIEPGPLYIPSSYVPSHHEYP
                     AMTFYSPAVMNYSIPSSVTNLEEGPGRQTTSPNMLWSTPGHLSPLAVHHQLSHLYAEP
                     QKSPWCEAKSLEHTLPVSRETLKRKVSGNHCASPVFGPSSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRREKCGYPLVPPQGNAEFQLHCAGKAKPSGGHVPPVPEFIISALSPFQLVLTLEAE
                     PPHVLISRPSVPFTEASMMMSLTKLADEELVHMISWAKKIPGFVELSLFDQVRLIESC
                     WLFVLMVGLMWRSIDHPGKLIFAPNLLLDRDEFGKCVEGILEVFDMLLATTSRFRELKL
                     QHKEYLCVKAMVLLNSQYDPLVTATQDAESSQKLAHLLNAVTDALVWVIAKSGFSSQQ
                     QSVRLANLLMILSHIRHAPAEKASHSLTSPGMEDGD"
BASE COUNT      378 a    410 c    411 g    366 t      1 others
ORIGIN

  Query Match             68.9%;  Score 1308.2;  DB 9;  Length 1566;
  Best Local Similarity   93.6%;  Pred. No. 4.1e-301;
  Matches 1364,  Conservative    0,  Mismatches   93;  Indels    0;  Gaps    0;

Qy          26 GACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATA 85
               | |||||||||||||| || ||||||| ||  |||||||||||||||||||||||||||
Db           1 GCCCTTTGTGCCTCTTCTCGCGAGGTGTTATCCTAGCTGTTATCTCAAGACATGGATATA 60

Qy          86 AAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTA 145
               |||||||||||||| |||||||||||||||| ||||||| |  ||||||||||||||
Db          61 AAAAACTCACCATCCAGCCTTAATTCTCCTTCTTCCTACAATTTTGGTCAATCCATCTTA 120

Qy         146 CCCCTCGAGCACCCCTCCATATACATACCTTCCTCCTATCTACACCCACCATCAATAT 205
               ||| |||||||||| | |||||||||||||||||||||||||||||||||||||||||
Db         121 CCCTTGGAGCACGGCCCTATATACATACCTTCCTCCTATGTAGAGAGCCACCATGAATAT 180

Qy         206 CCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTTACT 265
               || |||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db         181 CCGGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAGTGTTACT 240

Qy         266 AACTTGCAAGGTGGCCCTGGTCGGCAGACCACAAGCCCAAATATGCTGTGGTCAACACCT 325
               |||||||||| |||||||||||||| |||||||||||| || ||||| |||||||
Db         241 AACTTCCAAGAGGGCCTGGTCGGCAGATCACAAGCCCAAATATGCTGTGGTCAACACCT 300

Qy         326 GGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAA 385
               || |||||||||||||| |||||| ||||||||||||||||||||||||||||||||||
Db         301 GGCCACCTTTCTCCTTTAGCTGTCCATCACCAGTTATCACATCTGTATGCAGAACCTCAA 360

Qy         386 AAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACA 445
```

```
                ||||||||||||||||||||||||||| |||||||| |||||||||||| ||||||||||
Db        361 AAGAGTCCCTGGTGTGAAGCAAGATCACTAGAACATACCTTACCTGTAAGCAGAGAGACA 420

Qy        446 CTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAG 505
                |||||||||| |||||||||||| ||| ||||||||||||||||||| ||||||||
Db        421 CTGAAAACGAAAGTTAGTGGGAACCATTGTGCCAGCCCTGTTACTGGTCCAAGTTCAAAG 480

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
                |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||
Db        481 AGGGATGCTCACTTCTGTGCAGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 540

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 600

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 660

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
                |||||||||||||||| |||||||||||||| |||||||||||||||||||||||||
Db        661 TGCCGACTTCGGAAGTGCTACGAAGTGGGAATGGTAAAGTGTGGCTCCCGGAGAGAGAGA 720

Qy        746 TGTGGGTACCGCCTTCTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGC 805
                |||||||||||||||||||||||||| ||| |||||| |||||||||| || ||||||
Db        721 TGTGGGTACCGCCTTGTGCGGAGACAGGGAAATGCCGAGGAGCAGCTGCATTGCGCCGGC 780

Qy        806 AAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGACTGCGGGAGCTGCTGCTGGACGCCCTG 865
                |||||||||||||||| || |||| |||||||||||||||||||||||  |||||||
Db        781 AAGGCCAAGAGAAGTGGTGGTCACGTGCCCCGAGTGCGGGAGCTGCTGCTGACGCCCTG 840

Qy        866 AGCCCCGAGCAGCTAGTGCTCACCCTCCTGCAGCCTCAGCCGCCCCCATGTGCTGATCAGC 925
                |||||||||||||| |||||||||||||||||| || |||||||||||||||||||||
Db        841 AGCCCCGAGCAGCTGGTGCTCACCCTCCTGGAGGCCGAACCGCCCCCATGTGCTGATCAGC 900

Qy        926 CGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGAC 985
                |||||||| | |||||||| ||||||||||||||||||||||||||||||||||||||
Db        901 CGCCCCAGCGTGCCCTTCACTGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCCGAC 960

Qy        986 AAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGC 1045
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGC 1020

Qy       1046 CTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTG 1105
                ||||| |||||||||||||||||||||||| ||| ||||||||||| ||||||||||
Db       1021 CTGTTGGACCAAGTGCGGCTCTTGGAGAGCTGCTGGTTGGAGGTGTTAATGGTGGGGCTG 1080

Qy       1106 ATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGAC 1165
                |||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||
Db       1081 ATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAAATCTCATTCTGGAC 1140

Qy       1166 AGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACT 1225
                |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1141 AGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGAAGTCTTTGACATGCTCCTGGCAACT 1200

Qy       1226 ACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATG 1285
                |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db       1201 ACTTCAAGGTTTCGAGAGTTGAAACTCCAACACAAAGAGTATCTCTGTGTCAAGGCCATG 1260

Qy       1286 ATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGC 1345
                ||||||||||||||       ||||||||||||||| |||||||||||| ||||||
Db       1261 GTCCTGCTCAATTCCCAGTATGACCCTCTGGTCACAGCAACCCAGGATGCTGAGAGCAGC 1320

Qy       1346 CGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAG 1405
                | ||||||||| |||||||||| |||||||| || |||||||| |||||||| || |||
Db       1321 CAGAAGCTGGCCCACTTGCTGAATGCCGTGACTGACGCTTTGGTCTGGGTGATCGCTAAG 1380

Qy       1406 AGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCC 1465
                || ||| |||||||||||||||| ||  ||||||||||||||||||||||||||||||
Db       1381 AGTGGCTTCTCCTCCCAGCAGCAGTCGGTGCGCCTGGCTAACCTCCTGATGCTCCTGTCC 1440

Qy     * 1466 CACGTCAGGCATGCGAG 1482
                ||| ||||||||||||||
```

```
Db        1441 CACATCAGGCATGCGAG 1457


RESULT 19
BTA18017
LOCUS       BTA18017                1820 bp    mRNA    linear   MAM 09 SEP 1999
DEFINITION  Bos taurus mRNA for estrogen receptor beta.
ACCESSION   Y18017
VERSION     Y18017.1  GI:5823588
KEYWORDS    estrogen receptor beta.
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1820)
  AUTHORS   Walther,N., Lioutas,C., Tillmann,G. and Ivell,R.
  TITLE     Cloning of bovine estrogen receptor beta (ERbeta): expression of
            novel deleted isoforms in reproductive tissues
  JOURNAL   Mol. Cell. Endocrinol. 152 (1-2), 37-45 (1999)
  MEDLINE   99359179
   PUBMED   10432221
REFERENCE   2  (bases 1 to 1820)
  AUTHORS   Walther,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1998) N. Walther, Institute for Hormone and
            Fertility, Research, Grandweg 64, D-22529 Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .1820
                     /organism "Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_type "granulosa cells"
     gene            153. .1577
                     /gene="erbeta"
     CDS             153. .1577
                     /gene "erbeta"
                     /codon_start=1
                     /product "estrogen receptor betaa"
                     /protein_id="CAB53861.1"
                     /db_xref="GI:5823589"
                     /translation="MNYSIPDNSEDGPGRQTTSPNVLWPTPGHLSPLATHCQPSVLYA
                     EPQKSPWRETRSLEHTLPVNRETLKRKASGSSCASPATSPSSKRDAHFCAVCSDYASG
                     YHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVK
                     CGSPRERCGYRIVPRQRNSDEQLHCLSKTKPNGGPMTPVKFLLLSALSPEQLVLTLLE
                     AEPPHVLISRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLE
                     SCWLEVLMVGLMWRSIDHPGKLIFAPDLILDRDEGKCVEGILEIFDMLLATTSRFREL
                     KLQHKEYLCVKAMILLNSSMYPSATAPQEADSGRKLTHLLNAVTDALVWVIAKSGMSS
                     QQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHTLRGNKSL
                     VTGSERNLVEDSESKEGSQKPQAQ"
BASE COUNT      446 a    491 c    478 g    405 t
ORIGIN

  Query Match             66.2%;  Score 1256.4;  DB 4;  Length 1820;
  Best Local Similarity   86.1%;  Pred. No. 9.3e-289;
  Matches 1419;  Conservative    0;  Mismatches  216;  Indels   13;  Gaps    2;

Qy          84 TAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCT 143
               | ||||||||||||||||||||||| ||||||  ||||||  |||||||||||||||
Db           1 TCAAAAACTCACCATCTAGCCTTAACTCTCCTGTCTCCTATAACTGCGGTCAATCCATCC 60

Qy         144 TACCCCTGGAGCCAGGCCCCATCTATTTACCCTCTTCCTATGTAGACAGCCACCATGAAT 203
               |||||||||||  ||| |||| || ||||  || |||||||||| |||| ||||||||
Db          61 TACCCCTGGAGCCAGGCCCCATCTATTTACCCTCTTCCTATGTAGAGAGCCGCCATGAAT 120

Qy         204 ATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCA 263
               |  ||||| ||||||||||||||||||||||||||||||||||||| |||||         |
Db         121 ACTCAGCCGTGACATTCTATAGCCCTGCTGTGATGAATTACAGTATTCC--  ---  - CG 171

Qy         264 CTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACAC 323
               |||||  |||| |||||||||||| |||||||||| |||||||| |||||||||| |
Db         172 ATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGTCCAAATGTATTGTGGCCAACTC 231

Qy         324 CTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTC 383
```

```
Db        232 CTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCAGTTCTGTATGCAGAACCTC 291

Qy        384 AAAAGAGTCCCTGGTGTGAAGCAAGATCGCTACAACACACCTTACCTGTAAACAGAGAGA 443

Db        292 AAAAGAGTCCTTGGCGTGAAACCAGATCATTAGAACACACCTTACCTGTGAACAGGGAGA 351

Qy        444 CACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAA 503

Db        352 CACTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGCTACTAGTCCAAGTTCAA 411

Qy        504 AGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAG 563

Db        412 AGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCGGGATATCACTATGGAG 471

Qy        564 TCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATT 623

Db        472 TCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGTATTCAAGGACATAATGATT 531

Qy        624 ATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGG 683

Db        532 ATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGACGTAAAAGCTGTCAGG 591

Qy        684 CCTGCCGACTTCGGAAGTGTTACCAAGTGGGAATGGTGAAGTGTGGCTCCCGCACAGAGA 743

Db        592 CCTGTCGACTTCGGAAGTGCTATCAAGTTGGAATGGTGAAGTGTGGCTCTCGGAGAGAAA 651

Qy        744 GATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCG 803

Db        652 GGTGTGGGTACCGCATAGTGCGAAGGCAGAGAAATTCTGATGAGCAGCTGCACTGCCTGA 711

Qy        804 GCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCC 863

Db        712 GCAAAACCAAGAGGAAGGTGGAGCCATCACCCCACTGAAAGACCTCCTCCTCAGCGCCC 771

Qy        864 TGAGCCCCAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTCCTGATCA 923

Db        772 TGAGCCCACAGCAGCTGGTGCTTACGCTCCTGGAGGCGGACCCGCCCCCACCGTCCTATAA 831

Qy        924 GCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCCG 983

Db        832 GCCGCCCCAGCACGCCTTCACTGAGGCCTCCATGATGATGTCCCTCACCAAGCTGGCCG 891

Qy        984 ACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCA 1043

Db        892 ACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTCGTGGAGCTCA 951

Qy       1044 GCCTGTTCGACCAAGTGCTGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGC 1103

Db        952 GCCTGTACGACCAAGTGCTGCTTTTGCAGAGCTGCTGGTTGGAGGTGCTCATGGTGGGC 1011

Qy       1104 TGATGTGGCGCTCAATTGACCACCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGG 1163

Db       1012 TGATCTGGCGCTCCATCGACCACTCCTGGCAAGCTCATCTTTGCTCCAGATCTCATTCTGG 1071

Qy       1164 ACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAA 1223

Db       1072 ACAGGGATGAAGGGAAATGTGTTGAAGGAATTCTAGAAATCTTTGACATGCTCCTGGCAA 1131

Qy       1224 CTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCA 1283

Db       1132 CGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCA 1191

Qy       1284 TGATCCTGCTCAATTCCAGTATGTACCCTTCACCTACAGCACCCCAGCAGGCTGACAGTG 1343

Db       1192 TGATCCTCCTCAACTCCAGTATGTACCCTTCACCTACAGCACCCCAGCAGGCTGACAGTG 1251

Qy       1344 GCCGGAAGCTGGCTCACTTGCTGAAGCGCGTGACCGATGCTTTGGTTTGGGCTGAATTGCA 1403

Db       1252 GCCGGAAGCTGACTCACTTGCTGAAGCGCGTGACCGACGCTCTGGTCTGGGCTGAATTGCA 1311

Qy       1404 AGAGCGGCATCTCCTCCCAGCAGCAATCCATGGGCCTGGCTAACCTCCTGATGCTCCTGT 1463

Db       1312 AGAGTGGCATGTCCTCCCAGCAGCAGTCCATGCGCCTGGCTAACCTGCTGATGCTCCTGT 1371

Qy       1464 CCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAA 1523

Db       1372 CTCACGTCAGGCACGCCAGTAACAAGGGCATGGAACACCTGCTCAACATGAAGTGCAAAA 1431

Qy       1524 ATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGT 1583

Db       1432 ACGTGGTCCCCGTCTACGATCTGCTGCTGGAGATGCTGAATGCCCACACACTTCGCGGCA 1491

Qy       1584 GCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGG 1643

Db       1492 ACAAGTCTTTGGTCACAGGGTCTGAGCGCAACCTGGTGGAGCACAGTGAGAGCAAAGAGG 1551

Qy       1644 GCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCTGAGGTG----AACTGGCCCACAG 1699

Db       1552 GCTCCCAGAAACCCCAGGCTCAGTGACGCCCAGCCCCGAGCTGTGGTCAGACTGCAGAGA 1611

Qy       1700 AGGTCACAAGCTGAAGCGTGAACTCCAG 1727

Db       1612 AGGTCAGAGGCTGAAGCATGAACTCCAG 1639
```

```
RESULT 20
AF110402
LOCUS       AF110402               1611 bp    mRNA    linear   MAM 11-APR-1999
DEFINITION  Bos taurus breed Holstein estrogen receptor-beta mRNA, complete
            cds.
ACCESSION   AF110402
VERSION     AF110402.1  GI:4580595
KEYWORDS    .
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1611)
  AUTHORS   Rosenfeld,C.S., Yuan,X., Manikkam,M., Calder,M.D., Garverick,H.A.
            and Lubahn,D.B.
  TITLE     Cloning, sequencing, and localization of bovine estrogen
            receptor-beta within the ovarian follicle
  JOURNAL   Biol. Reprod. 60 (3), 691-697 (1999)
  MEDLINE   99150196
   PUBMED   10026117
REFERENCE   2  (bases 1 to 1611)
  AUTHORS   Yuan,X., Rosenfeld,C.S., Manikkam,M., Calder,M.D., Garverick,H.A.
            and Lubahn,D.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Biochemistry, University of
            Missouri-Columbia, 163A Animal Sciences Research Center, 920 E.
            Campus Drive, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1..1611
                     /organism="Bos taurus"
                     /strain="breed Holstein"
                     /db_xref="taxon:9913"
                     /tissue_type="ovarian follicle"
                     /note="RT-PCR"
     CDS             28..1611
                     /function="steroid receptor"
                     /codon_start=1
                     /product="estrogen receptor-beta"
                     /protein_id="AAD24432.1"
                     /db_xref="GI:4580596"
                     /translation="MDVKNSPSSLNSPVSYNCQQSTLPLEPGPIYLPSSYVPSRHEYS
                     AVTEYSPAVMNYSIPNNSEDGPGKQTTSPNVLWPTPGHLSPLATHCQPSVLYAEPQKS
                     PWPETRSLEHTLPVNRETLKRKASGSSCASPATSPSSKRDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSR
                     ERCGYRIVRHQRNSDEQLHCLSKTKPNGGPMTRVKFLILSALSPFQLVLTLLEAEPPH
                     VLISPEPTPEPASMMMSLTFLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLILDRDEGKCVEGILEIFDMLLATTSRFRELKLQHK
                     EYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGMSSQQQSM
                     RLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHTLHGNKSSVTGSE
```

```
                     RNLVEDSESKEGSQKPQAQ"
BASE COUNT      405 a    428 c    416 g    362 t
ORIGIN

  Query Match             66.2%;  Score 1256.2;  DB 4;  Length 1611;
  Best Local Similarity   87.1%;  Pred. No. 1e-288;
  Matches 1394;  Conservative    0;  Mismatches  198;  Indels    9;  Gaps    1;

Qy          69 CTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACT 128
               |||||||||||||| | ||||||||||||||||||||||| ||||||  |||||| ||||
Db          20 CTCAAGACATGGATGTCAAAAACTCACCATCTAGCCTTAACTCTCCTGTCTCCTATAACT 79

Qy         129 GCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAG 188
               || |||||||||| |||||||||||  ||| |||| ||  |||| || ||||||||||
Db          80 GCGGTCAATCCATCCTACCCCTGGAGCCAGGCCCCATCTATTTACCCTCTTCCTATGTAG 139

Qy         189 ACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCA 248
               | |||| |||||||||  |||| |||||||||||||||||||||||||||||||||| |
Db         140 AGAGCCGCCATGAATACTCAGCCGTGACATTCTATAGCCCTGCTGTGATGAATTACAGTA 199

Qy         249 TTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATG 308
               ||||          || ||||  |||| ||||||||||| ||||||||||||||||||||
Db         200 TTCC    ----  -CAATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGCCCAAATG 250

Qy         309 TGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATC 368
               | ||||||||||| |||||  |||| ||||||||||    |||||  ||||    |||  ||
Db         251 TATTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCAGTTC 310

Qy         369 TGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTAC 428
               ||||||| |||||||||||||||| ||| |||||  | |||||  |||||||||||||||
Db         311 TGTATGCAGAACCTCAAAAGAGTCCTTGGCGTGAAACCAGATCATTAGAACACACCTTAC 370

Qy         429 CTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTA 488
               |||| ||||| |||||| |||||||||||| |||||||| | |||| ||||||||||  ||
Db         371 CTGTGAACAGGGAGACATTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGCTA 430

Qy         489 CTGGTCCAGGTTCAAAGAGGGATGCCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGC 548
               || ||||| |||||||||||||||| ||||||||  |||||||||| |||||  || ||||
Db         431 CTAGTCCAAGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCGC 490

Qy         549 GATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTC 608
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||| ||||
Db         491 GATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGTATTC 550

Qy         609 AAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGC 668
               |||||||||||||||||||||||||||||||||||||||||||||| || ||||| |||| |
Db         551 AAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGAC 610

Qy         669 GCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTG 728
               | || ||||| |||||||| |||||||||||||| || || || ||||||||||||||||
Db         611 GTAAAAGCTGTCAGGCCTGTCGACTTCGGAAGTGCTATGAGGTTGGAATGGTGAAGTGTG 670

Qy         729 GCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGC 788
               |||| |||||||| || |||||||||||| | ||||| || ||||||| | | || ||||
Db         671 GCTCTCGGAGAGAAAGGTGTGGGTACCGCATAGTGCGAAGGCAGAGAAATTCTGATGAGC 730

Qy         789 AGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGC 848
               ||||||||||     ||||  ||||||| |  || || | |  | ||||||||   ||||
Db         731 AGCTGCACTGCCTGAGCAAAACCAAGAGGAACGGTGGACCCATGACCCGAGTGAAAGAGC 790

Qy         849 TGCTGCTGGACGCCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGC 908
               ||||||||  |||||||||||| ||||||||  |||||  || ||||||||||| |||||||
Db         791 TGCTGCTGAGCGCCCTGAGCCCAGAGCAGCTGGTGCTTACGCTCCTGGAGGCCGAGCCGC 850

Qy         909 CCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCC 968
               |||| |||||  || ||||||||||  |||||||||| ||||||||||||||||||||||
Db         851 CCCACGTGCTCATAAGCCGCCCCAGCACGCCCTTCACTGAGGCCTCCATGATGATGTCCC 910

Qy         969 TGACCAAGTTGGCCGACAAGGAGTTCGTACACATGATCAGCTGGGCCAAGAAGATTCCCG 1028
               | ||||| |||||||||||||  |||||||||||||||||||||||||||||||||| |
Db         911 TCACCAAGCTGGCCGACAAGGAACTCGTACACATGATCAGCTGGGCCAAGAAGATTCCCG 970

Qy        1029 GCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGC 1088
               |||| ||||||||||||||| |||||||||||||| |||||||||| ||| ||||||
Db         971 GCTTCGTGGAGCTCAGCCTGTACGACCAAGTGCGGCTTTTGGAGAGCTGCTGGTTGGAGG 1030

Qy        1089 TGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTC 1148
               || | ||| |||||||||||||||||| || |||||||| ||||||||||||||||||
Db        1031 TGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCTGGCAAGCTCATCTTTGCTC 1090

Qy        1149 CAGATCTTGTTCTTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTG 1208
               |||| ||  |||||||||||||||| |||||||| || |||||||||| ||||||||||
Db        1091 CAGACCTCATTCTGGACAGGGATGAAGGGAAATGTGTTGAAGGAATTCTAGAAATCTTTG 1150

Qy        1209 ACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATC 1268
               |||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||
Db        1151 ACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATATC 1210

Qy        1269 TCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCC 1328
               |||||||||||||||||||| |||| ||||||||||||||   | ||||| |||
Db        1211 TCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTTCAGCTACAGCACCCC 1270

Qy        1329 AGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGG 1388
               |||| ||||||||  |||||||||| |||| ||||||| || |||| || ||| |||
Db        1271 AGGAGGCTGACAGTGGCCGGAAGCTGACTCACCTGCTGAATGCTGTGACGGACGCTCTGG 1330

Qy        1389 TTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACC 1448
               | ||||||||||||||| ||||| |||||||||||||| ||||||||||||||||||||
Db        1331 TCTGGGTGATTGCCAAGAGTGGCATGTCCTCCCAGCAGCAGTCCATGCGCCTGGCTAACC 1390

Qy        1449 TCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCA 1508
               | ||||||||||||| |||||||||| || |||||||||||||||||| |||||||
Db        1391 TGCTGATGCTCCTGTCTCACGTCAGGCACGCCAGTAACAAGGGCATGGAACACCTGCTCA 1450

Qy        1509 ACATGAAGTGCCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCC 1568
               |||||||||||||||| |||||||| || || || ||||||||||||||||||||||||
Db        1451 ACATGAAGTGCAAAAACGTGGTCCCCGTCTACGATCTGCTGCTGGAGATGCTGAATGCCC 1510

Qy        1569 ACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACA 1628
               ||   ||||||||   |||||| |   |||| |||| ||| || || ||  |||||||
Db        1511 ACACACTTCGCGGCAACAAGTCTTTGGTCACAGGGTCTGAGCGCAACCTGGTGGAGGACA 1570

Qy        1629 GTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
               || | |||||||||||||||||||| || ||| |||||||
Db        1571 GTGAGAGCAAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1611


RESULT 21
AF177936
LOCUS       AF177936                1584 bp    mRNA    linear   MAM 16-JUL 2001
DEFINITION  Ovis aries estrogen receptor beta mRNA, complete cds.
ACCESSION   AF177936
VERSION     AF177936.1  GI:5911382
KEYWORDS    .
SOURCE      Ovis aries.
  ORGANISM  Ovis aries
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 1584)
  AUTHORS   Cardenas,H., Burke,K.A., Bigsby,R.M., Pope,W.F. and Nephew,K.P.
  TITLE     Estrogen receptor beta in the sheep ovary during the estrous cycle
            and early pregnancy
  JOURNAL   Biol. Reprod. 65 (1), 128-134 (2001)
  MEDLINE   21313442
   PUBMED   11420232
REFERENCE   2  (bases 1 to 1584)
  AUTHORS   Cardenas,H., Bigsby,R.M., Pope,W.F. and Nephew,K.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1999) Obstetrics and Gynecology, Indiana
            University, 1001 Walnut St., Indianapolis, IN 46202, USA
FEATURES             Location/Qualifiers
     source          1. .1584
                     /organism="ovis aries"
                     /db_xref="taxon:9940"
                     /tissue_type "ovary"
```

```
     CDS             1. .1584
                     /note="steroid receptor"
                     /codon_start 1
                     /product="estrogen receptor beta"
                     /protein_id="AAD55772.1"
                     /db_xref "GI:5911383"
                     /translation="MDVKNSPSSLNSPVSYNCQQSILPLEPCPIYLPSSYVFSRPEYS
                     AVIFYSPAVMNYSIPNNSEDGPGRQTTSPNVLWPTPGHLSPLAIHCQSSLLYAEPQKS
                     PWCEIRSLEHTFPVNRETLKRKASGSSCASPVSSPSSKRDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRK
                     ERCGYRIVRRQRNSDEQLHCLSKTKRNGAPMTRVKELLLSALSPEQLVLTLLEAEPPH
                     VLMSRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDEDEGKCVEGILEIFDMLLATTSRFRELKLQHK
                     EYLCVKAMILLNSSMYPSATASQEADSGRKLTHLLNAVTDALVWVIAKSGMSSQQQSM
                     RLANLLMILSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSE
                     RNLAEDSESKEGSQKPQAQ"
BASE COUNT      397 a    420 c    414 g    353 t
ORIGIN

  Query Match             66.1%;  Score 1254.6;  DB 4;  Length 1584;
  Best Local Similarity   87.3%;  Pred. No. 2.5e-288;
  Matches 1390;  Conservative    0;  Mismatches  194;  Indels    9;  Gaps    1;

Qy         77 ATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAA 136
              |||||| | ||||||||||||||||||||||| |||||| |||||| |||||| |||||
Db          1 ATGGATGTCAAAAACTCACCATCTAGCCTTAACTCTCCTGTCTCCTATAACTGCGGTCAA 60

Qy        137 TCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCAC 196
              ||||| |||||||||||  ||| |||||||  |||| || ||||||||||| |||| |
Db         61 TCCATTCTACCCCTGGAGCCAGGCCCCATATATTTACCCTCTTCCTATGTAGAGAGCCGC 120

Qy        197 CATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGC 256
              ||||||||  ||||| ||||||||||||||||||||||||||||||||| |||||
Db        121 CATGAATACTCAGCCGTGACATTCTATAGCCCTGCTGTGATGAATTACAGTATTCC---- 176

Qy        257 AATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGG 316
              || |||||  |||| |||||||||||| ||||||||||||||||||| ||||||
Db        177 -----CAATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGCCCAAATGTATTGTGG 231

Qy        317 CCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCG 376
              ||||| |||| |||| ||||||||||  |||| |||||  |||| |||||||||
Db        232 CCAACTCCTGGACACCTCTCCCTTTAGCCATCCATTGCCAGTCGTCACTTCTGTATGCA 291

Qy        377 GAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAAC 436
              |||||||||||||||| |||||||| | |||||  |||||||||||| ||||| |||
Db        292 GAACCTCAAAAGAGTCCTTGGTGTGAAACCAGATCATTAGAACACACCTTCCCTGTGAAC 351

Qy        437 AGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCA 496
              || |||||| |||||||||| | |||||||| | |||| ||||||||||| || ||||
Db        352 AGGGAGACATTGAAAAGGAAAGCTAGTGGGAGCAGTTGTGCCAGCCCTGTTTCTAGTCCG 411

Qy        497 GGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCAC 556
              |||||||||||||||| |||||||| |||||||||||||||| || ||||||||||||
Db        412 AGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGCGATTATGCGTCGGGATATCAC 471

Qy        557 TATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACAT 616
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        472 TATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACAT 531

Qy        617 AATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGC 676
              |||||||||||||||||||||||||||||||||||| ||||| |||| || || |||
Db        532 AATGATTATATTTGTCCAGCTACAAATCAGTGTACCATACATAAGAACCGACGTAAAACC 591

Qy        677 TGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGG 736
              || |||||||| |||| |||||||| || || || ||||||||||||||||||||||||
Db        592 TGTCAGGCCTGTCGACTGCGGAAGTGCTATGAGGTCGGAATGGTGAAGTGTGGCTCCCGG 651

Qy        737 AGAGAGAGATGTGGGTACCGCTTGTGCGGAGACAGACAAGTGCCGACCAGCAGCTGCAC 796
              |||||  || |||||||||||| | ||||| || |||||| | | || ||||||||||||
Db        652 AGAGAAAGGTGTGGGTACCGCATAGTGCGAAGGCAGACAAATTCTGATCACCAGCTGCAC 711

Qy        797 TGTGCCGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTG 856
              ||     ||||  ||||||| |  || |  | |  | ||||||||  ||||||||||||

Db        712 TGCCTGAGCAAAACCAAGAGGAACGGTGCACCCATGACCCGAGTGAAGGAGCTGCTGCTG 771

Qy        857 GACGGCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTG 916
              |||||||||| |||||||| |||| || ||||||||||  |||||||||| |||
Db        772 AGTGCCCTGAGCCCAGAGCAGCTGGTGCTTACTCTCCTGGAGGCCGAGCCGCCCCACGTG 831

Qy        917 CTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAG 976
              || || |||||||||| |||||||||| |||||||||||||||||||||| |||||||
Db        832 CTCATGAGCCGCCCCAGCGCGCCCTTCACTGAGGCCTCCATGATGATGTCCTTGACCAAG 891

Qy        977 TTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTG 1036
              ||||||||||||  |||||||||||||||||||||||||||||||||| |||| |||
Db        892 CTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTCGTG 951

Qy       1037 GAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATG 1096
              |||||||||||| |||||||||||||||||||||||| ||| ||||||| | |||
Db        952 GAGCTCAGCCTGTACGACCAAGTGCGGCTCTTGGAGAGCTGCTGGTTGGAGGTGCTCATG 1011

Qy       1097 ATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTT 1156
              |||||||||||||||| || |||||||||||||||||||||||||||||| ||
Db       1012 GTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTC 1071

Qy       1157 GTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTC 1216
              |||||||||||||||| |||||||| || |||||||||| ||||||||||||||||||
Db       1072 GTTCTGGACAGGGATGAAGGAAATGTGTTGAAGGAATTCTAGAAATCTTTGACATGCTC 1131

Qy       1217 CTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTC 1276
              ||||||| |||||||||||| |||||||||||||||||||||||||||||||||||||
Db       1132 CTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTC 1191

Qy       1277 AAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT 1336
              |||||||||||||| ||||| |||||||||||||||  |  |||||  |||||| |||
Db       1192 AAGGCCATGATCCTGCTCAACTCCAGTATGTACCCTTCAGCTACAGCATCCCAGGAGGCT 1251

Qy       1337 GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTG 1396
              |||||  |||||||||| ||||| ||||||| |||||||| || ||| |||| ||||||
Db       1252 GACAGTGGCCGGAAGCTGACTCACCTGCTGAATGCCGTGACGGACGCTCTGGTCTGGGTG 1311

Qy       1397 ATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTGCTGATG 1456
              |||||||||||||||| |||||||||| || ||||||||||||||||||| ||||||
Db       1312 ATTGCCAAGAGCGGCATGTCCTCCCAGCAACAGTCCATGCGCCTGGCTAACCTGCTGATG 1371

Qy       1457 CTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAG 1516
              |||||||| || |||||||||| ||||||||||||||||||||||||| ||||||||||||
Db       1372 CTCCTGTCTCATGTCAGGCATGCCAGTAACAAGGGCATGGAACACCTGCTCAACATGAAG 1431

Qy       1517 TGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTT 1576
              |||||||||||||| || || || || ||||||||||||||||||||||||||  ||||
Db       1432 TGCAAAAATGTGGTGCCGTCTACGATCTGCTGCTGGAGATGCTGAATGCCCACGTGCTT 1491

Qy       1577 CGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGC 1636
              ||| |   |||| | |   |||| |||| ||| ||| || ||| |||||||| | |||
Db       1492 CGCAGCAACAAGCCTTTGGTCACAAGGGTTGAGCGCAACCCTGGGGGAGGACAGCCAGAGC 1551

Qy       1637 AAAGAGGGCTCCCAGAAGCCCACAGTCTCAGTGA 1669
              |||||||||||||||| || ||| ||||||||
Db       1552 AAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1584


RESULT 22
A65105
LOCUS       A65105                  1251 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 2 from Patent EP0798378.
ACCESSION   A65105
VERSION     A65105.1  GI:4530974
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1251)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 2 01-OCT-1997;
```

```
            AKZO NOBEL NV (NL)
COMMENT     Other publication CA 2200423 19970926
            Other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1. .1251
                     /organism "unidentified"
                     /db_xref "taxon:32644"
BASE COUNT       297 a     324 c     354 g     276 t
ORIGIN

  Query Match             65.7%;  Score 1247;  DB 6;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-286;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy          296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy          356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy          416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy          476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy          536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy          596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy          656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy          716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy          776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy          836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy          896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 GAGGCTGAGCCGCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy          956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy         1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy         1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy         1136 CTCATCTTTGCTCCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGCAATT 1195
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 CTCATCTTTGCTCCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGCAATT 960

Qy         1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy         1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy         1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy         1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy         1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
                |||||||||||||||||||||||||||||||||||||||||||||||
Db         1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 23
AX323056
LOCUS       AX323056                1251 bp    DNA     linear   PAT 07-JAN 2002
DEFINITION  Sequence 2 from Patent EP1162264.
ACCESSION   AX323056
VERSION     AX323056.1  GI:18094943
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 2 12-DEC 2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1251
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
BASE COUNT       297 a     324 c     354 g     276 t
ORIGIN

  Query Match             65.7%;  Score 1247;  DB 6;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-286;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy          296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy          356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy          416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy          476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
```

```
Qy        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAAATC 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAAATC 420

Qy        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        716 ATGCTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCC 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCC 600

Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCGGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCCTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCCTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCCTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCCTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247

RESULT 24
A65123
LOCUS       A65123                  1257 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 20 from Patent EP0798378
ACCESSION   A65123
VERSION     A65123.1  GI:4530988
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1257)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 20 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     Other publication CA 2200423 19970926
            Other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1..1257
                     /organism="unidentified"
                     /db_xref "taxon:32644"
BASE COUNT      297 a    327 c    355 g    278 t
ORIGIN

  Query Match           65.7%;  Score 1247;  DB 6;  Length 1257;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-286;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         236 ATGAATTACACCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy         356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy         416 CAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGCTTACTGGGAACCCTTGC 475
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGCTTACTGGGAACCCTTGC 240

Qy         476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy         536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy         596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAAATC 655
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAAATC 420

Qy         656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy         716 ATGCTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy         776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCC 835
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCC 600

Qy         836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy         896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy         956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 1015
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy        1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
```

```
                ||||||
```

```
LOCUS       RNAJ2602                1650 bp    DNA     linear   ROD 18-APR-1998
DEFINITION  Rattus norvegicus mRNA for estrogen receptor beta (clone
            pcDNA.ERbeta).
ACCESSION   AJ002602
VERSION     AJ002602.1  GI:3077648
KEYWORDS    estrogen receptor beta.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) Aldridge T.C., Molecular Endocrinology
            Group, ZENECA Central Toxicology Laboratory, Alderley Park,
            Macclesfield, Cheshire SK10 4TJ, ENGLAND
  REMARK    revised by [3]
REFERENCE   2  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Tissue-specific responses to estrogen: an explanation based on
            differential activation of multiple estrogen receptors with
            different estrogen response elements
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1997) Aldridge T.C., Molecular Endocrinology
            Group, ZENECA Central Toxicology Laboratory, Alderley Park,
            Macclesfield, Cheshire SK10 4TJ, ENGLAND
COMMENT     On Apr 23, 1998 this sequence version replaced gi:2632166.
FEATURES             Location/Qualifiers
     source          1. .1650
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone="pcDNA.ERbeta"
                     /tissue_type="Prostate"
     gene            1. .1650
                     /gene="ERbeta"
     CDS             1. .1650
                     /gene="ERbeta"
                     /note="The clone was isolated by RT-PCR from total RNA"
                     /codon_start=1
                     /evidence=experimental
                     /product="estrogen receptor beta"
                     /protein_id="CAA05631.1"
                     /db_xref="GI:3077649"
                     /db_xref="SPTREMBL:O35784"
                     /translation="MSICTSSHKFFSQLRPSEDMEIKNSPSSISSPASYNCSQSTIPI
                     EHGPIYIPSSYVDNRHEYSAMTFYSPAVMNYSVPGSTSNLDGGPVRQSTSPNVLWPTS
                     GHLSPLATHCQSSLLYAFPQKSPWCEARSLFETLPVNRFTLKRKLSGSSCASPVTSPN
                     AKRDAHFCAVCSDYASGYHYGVWPCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRK
                     SCQACRLRKCYEVGMVKGGIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWP
                     LLLSTLSPEQLVTLIEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKK
                     IPGFVELSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGI
                     LEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLTHLLN
                     AVTDALVWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYD
                     LLLEMLNAHVLRGYKPSISGSEGSSTEDSKNKESSQNLQSQ"
BASE COUNT      415 a    443 c    432 g    360 t
ORIGIN

  Query Match           65.4%;  Score 1240.2;  DB 10;  Length 1650;
  Best Local Similarity   85.0%;  Pred. No. 6.7e-285;
  Matches 1399;  Conservative    0;  Mismatches  243;  Indels    3;  Gaps    1;

Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTA---TCTCAAGACATGGATAT 84
              | | ||| ||||||||  ||||| ||||||||||||||  |   ||| ||||||||| ||
Db          6 CATCTGTACCTCTTCTCACAAGGAGTTTTCTCAGCTGCGACCCTCTGAAGACATGGAGAT 65

Qy         85 AAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTT 144
               |||||||||||| || ||||||| ||| ||| | ||||| ||||| || || || ||||||| |
Db         66 CAAAAACTCACCGTCGAGCCTTAGTTCCCCTGCTTCCTATAACTGTAGCCAGTCCATCCT 125

Qy        145 ACCCCTGGAGCACGGCTCCATATACATAGCTTCCTCCTATGTAGACAGTCAGCATGAATA 204
              |||||||||||||||| |||| ||||| |||||||||||| ||||||| || |||||| ||
Db        126 ACCCCTGGAGCACGGCCCCATCTACATCCCTTCCTCCTACGTAGACAACCGCCATGAGTA 185

Qy        205 TCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCAC 264
              | |||| |||||||||| || ||||||||||||| ||||  ||||| ||| || ||
Db        186 TTCAGCTATGACATTCTACAGTCGTGCTGTGATGAACTACAGTGTTCCCGGCAGCACCAG 245

Qy        265 TAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACC 324
              |||| |||| |||||||||  || |||| |||||||||||||||| | |||||||  |
Db        246 TAACCTGGACGGTGGGCCTGTCCGACAGAGCACAAGCCCAAATGTGCTATGGCCAACTTC 305

Qy        325 TGGGCACCTTTCTCCTTTAGTGGTCCATGGCCAGTTATCACATCTGTATGCGGAACCTCA 384
              ||||||||| |||||||||| |  |||| |||| | ||| |  || ||||| |||||||
Db        306 TGGGCACCTGTCTCCTTTAGCGACCCATTGCCAATCATCGCTCCTCTATGCAGAACCTCA 365

Qy        385 AAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGAC 444
              |||||||| |||||||||||||||| ||||| ||||||||||||||||||||||||||||
Db        366 AAAGAGTCCTTGGTGTGAAGCAAGATCACTAGAGCACACCTTACCTGTAAACAGAGAGAC 425

Qy        445 ACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAA 504
              |||||| |||||| |||||||| | |||| ||||||||||||||| |||||    ||||
Db        426 ACTGAAGAGGAAGCTTAGTGGGAGCAGTTGTGCCAGCCCTGTTACTAGTCCAAACGCAAA 485

Qy        505 GAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGT 564
              |||||||||||||||||||| |||||||||||||| |||||  || ||||| || || ||
Db        486 GAGGGATGCTCACTTCTGCGCCGTCTGCAGCGATTATGCATCTGGGTATCATTACGGCGT 545

Qy        565 CTGCTCGTCTGAAGGATCTAAGCCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTA 624
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TTGCCATGTGAAGGATGTAAGCCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTA 605

Qy        625 TATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGC 684
              ||| |||||||| || |||||||||| || || || |||||||| || ||||||||||
Db        606 TATCTGTCCAGCCACGAATCAGTGTACCATAGACAAGAACCGGCGTAAAAGCTGCCAGGC 665

Qy        685 CTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGCTGAAGTATGGCTCCCTCCCGAGACAC 744
              |||||||||||| |||||||| |||| |||||||| |||||||| ||| |||||||  |
Db        666 CTGCCGACTTGCAAGTGTTATGAAGTAGGAATGCTCAAGTGTGGATCCAGGAGACAACC 725

Qy        745 ATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGAGTAGCAGCTGCACTGTGCCGG 804
              |||||||||||  | |||||||| |||||||| |   ||||||| | |||||     |
Db        726 GTGTGGGTACCGCTATAGTGCGGAGGCAGAGAAGTTCTAGCGAGCAGGTACACTGCCTGAG 785

Qy        805 CAAGGCCAAGAGAAGTGGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCT 864
              ||| ||||||||||  || || ||||| |||| |||  |||||| ||||||  | || |
Db        786 CAAAGCCAAGAGAAACGGTGGGCAGGCACCCCGGGTGAAGGAGCTACTGCTGAGCACCTT 845

Qy        865 GAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAG 924
              ||| || |||| || |||||||||||||||| |||| || ||| |||||||| | ||
Db        846 GAGTCCAGAGCAACTGGTGCTCACCCTCCTGGAAGCTGAACCACCCAATGTGCTGGTGAG 905

Qy        925 CCGCCCCAGTGCGGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGA 984
              ||| |||||  |||||||||||||||||||||||||||||||||| || ||| ||| ||
Db        906 CCGTCCCAGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTCACTAAGCTGGCGGA 965

Qy        985 CAAGGAGTTGGTGCACATGATTGGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAG 1044
              |||||  |||| |||||||| |||||||||||| || || |||||||||||||||||||
Db        966 CAACGAACTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAG 1025

Qy       1045 CCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCT 1104
              |||||| |||||||| |||||||| || ||||| |||||||||||| ||||| |||| ||
Db       1026 CCTGTTGGACCAAGTCCGGCTCTTAGAAAGCTGCTGGATGGAGGTGCTAATGGTGGGACT 1085

Qy       1105 GATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGA 1164
              |||||||||| || |||||||||||||||||||||||| || ||||| || || ||||||||
Db       1086 GATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATTTTCGCTCCCGACCTCGTTCTGGA 1145

Qy       1165 CAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAAC 1224
              |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||| ||
Db       1146 CAGGGATGAGGGGAAGTGCGTAGAAGGGATTCTGGAAATCTTTGACATGCTCCTGGCGAC 1205
```

```
Qy       1225 TACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCAT 1284
             || |||||||| || ||||||||||||||| ||||| || ||||||||||| ||||||||
Db       1206 GACGTCAAGGTTCCGTGAGTTAAAACTCCAGCACAAGGAGTATCTCTGTGTGAAGGCCAT 1265

Qy       1285 GATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAG 1344
             |||||| ||||| |||||||||||||| ||| | || | |||||| || || || ||
Db       1266 GATCCTCCTCAACTCCAGTATGTACCCCTTGGCTTCTGCAAACCAGGAGGCAGAAAGTAG 1325

Qy       1345 CCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAA 1404
             |||||||||| | ||| | |||||||| ||||| ||||| |||| ||||||||||| ||
Db       1326 CCGGAAGCTGACACACCTACTGAACGCGGTGACAGATGCCCTGGTCTGGGTGATTGCGAA 1385

Qy       1405 GAGCGGCATCTCCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTC 1464
             ||| || |||||||||||||||| || | || ||||| ||||||||||||| || ||
Db       1386 GAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCGACTGGCCAACCTCCTGATGCTTCTTTC 1445

Qy       1465 CCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAA 1524
             ||||||||||     |||||||||||||||||||||||||||| ||||||||||||||||
Db       1446 TCACGTCAGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAA 1505

Qy       1525 TGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTG 1584
             |||||||| |||||||||||||||||||||||||||||||||| ||| ||||| ||||
Db       1506 TGTGGTCCCGGTGTATGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTA 1565

Qy       1585 CAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGG 1644
             |||| |||| ||| |||||| |||||||| | ||||||||||||| | |||||||
Db       1566 CAAGCCCTCAATCTCGGGGTCTGAGTGCAGCTCAACAGAGGACAGTAAGAACAAAGAGAG 1625

Qy       1645 CTCCCAGAACCCACAGTCTCAGTGA 1669
             |||||||||| ||||||||||||
Db       1626 CTCCCAGAACCTACAGTCTCAGTGA 1650


RESULT 27
RNU57439
LOCUS       RNU57439                2555 bp    mRNA    linear   ROD 12-JUN-1996
DEFINITION  Rattus norvegicus estrogen receptor beta mRNA, complete cds.
ACCESSION   U57439
VERSION     U57439.1  GI:1373280
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 8 to 2555)
  AUTHORS   Kuiper,G.G., Enmark,E., Pelto Huikko,M., Nilsson,S. and
            Gustafsson,J.A.
  TITLE     Cloning of a novel receptor expressed in rat prostate and ovary
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (12), 5925-5930 (1996)
  MEDLINE   96244066
   PUBMED   8650195
REFERENCE   2  (bases 1 to 2555)
  AUTHORS   Enmark,E.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03 MAY 1996) Eva L. Enmark, Center for Biotechnology,
            NOVUM, Karolinska, Institute, NOVUM, Huddinge, S-141 57, Sweden
FEATURES             Location/Qualifiers
     source          1..2555
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /sex="male"
                     /tissue_type="prostate"
     CDS             417..1874
                     /note="nuclear receptor superfamily member"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAC52602.1"
                     /db_xref="GI:1373281"
                     /translation="MTFYSPAVMNYSVPQSTSNLDGGPVRLSTSPNVLWPTSGHLSPL
                     ATHCQSSLLYAEPQKSPWCEARSLEHTLPVNRETLKRKLSGSSCASPVTSPNAKRDAH
                     FCPVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LKKCYEVGMVKGGSRKERGGYRQIRHKRSSSEQVHCLSKAKRNGGHAPKVKELLLSTL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLHDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLFHLLNAVTDAL
                     VWVIAKSGISSQQQSVRLANLLMILSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHTLRGYKSSISGSECSSTEDSKNKESSQNLQSQ"
     misc_feature    726..1923
                     /note="encodes DNA binding domain"
BASE COUNT      630 a    652 c    660 g    613 t
ORIGIN

  Query Match             65.8%;  Score 1239.6;  DB 10;  Length 2555;
  Best Local Similarity   83.6%;  Pred. No. 9.7e-285;
  Matches 1454;  Conservative    0;  Mismatches  279;  Indels    7;  Gaps    4;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              || || ||| |||||||||| ||      | | ||| |||||||| ||||| |||||||||
Db        206 CATGAGTCTCTGAGAACATAAT--GTCCATCTGTACCTCTTCTCACAAGGAGTTTTCTCA 263

Qy         61 GCTGTTA---TCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTC 117
              ||||  |   ||| |||||||||| || |||||||||| || ||||||| ||| ||| |
Db        264 GCTGCGACCCTCTGAAGACATGGAGATCAAAAACTCACCGTCGAGCCTTAGTTC-CCTGC 322

Qy        118 CTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTC 177
              ||||| ||||| || || ||||| |||||||||||||||| ||| |||| |||||
Db        323 TTCCTATAACTGTAGCCAGTCCATCCTACCCCTGGAGCACGGCCCCATCTACATCCCTTC 382

Qy        178 CTCCTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGAT 237
              |||||| |||||| || |||||| ||| |||| |||||||||| || ||||||||||
Db        383 CTCCTACGTAGACAACCGCCATGAGTATTCAGCTATGACATTCTACAGTCCTGCTGTGAT 442

Qy        238 GAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCAC 297
              ||| ||||| ||||| ||| || || |||| |||| |||||||||| || | || |||
Db        443 GAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGCAC 502

Qy        298 AAGCCCAAATGTGTTGTGGCCAACACCTGGCCACCTTTCTCCTTTAGTGGTCCATCGCCA 357
              |||||||||||| | |||||||  |||||||| |||||||| | ||| ||||
Db        503 AAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGCCA 562

Qy        358 GTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGA 417
              | ||| |  || ||||| |||||||||||||||| |||||||||||||||| |||||
Db        563 ATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTAGA 622

Qy        418 ACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGC 477
              |||||||||||||||||||||||||||||||| |||||| |||||||| | |||| ||
Db        623 GCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCACTTCTGC 682

Qy        478 CAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGA 537
              |||||||||||| |||||    |||||||||||||||||||||| | ||||||||||
Db        683 CAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGCGA 742

Qy        538 TTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAA 597
              ||| ||||| || ||||| || || || |||| ||||||||||||||||||||||||||
Db        743 TTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTTAA 802

Qy        598 AAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGA 657
              |||||||||||||||||||||||||||||||| |||||||| || |||||||||| || ||
Db        803 AAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATAGA 862

Qy        658 TAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAAT 717
              || |||||||| || ||||||||||||||||||||||| |||||||| ||||| |||||
Db        863 CAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGAAT 922

Qy        718 GGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAG 777
              ||| |||||||| ||| |||||||  | ||||||||||  | ||||||| |||||||
Db        923 GGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGAAG 982

Qy        778 TGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCG 837
              | |   ||||||| | |||||     |||| ||||||||||  || || || || |||||
Db        983 TTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCCCG 1042

Qy        838 AGTGCGGCAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGGTTCACCCTCCTGGA 897
                ||| |||||| ||||||  | || |||| || ||||| || ||||||||||||||||
```

```
Db       1043 GGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTGCTGGA 1102

Qy        898 GGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCAT 957
              ||||| || ||| |||||||| | ||||| |||||   ||||||||||||||||||||||
Db       1103 AGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCCAT 1162

Qy        958 GATGATGTCCCTGACCAAGTTGGCCGACAAGCACTTGCTACACATGATCAGCTGGCCCAA 1017
              |||||||||||| || ||| |||| |||||||  |||| |||||||  ||||||||||
Db       1163 GATGATGTCCCTCACTAAGCTGGCCGACAAGAACTTGGTGCACATGATTCGCTGGCCCAA 1222

Qy       1018 GAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCCACCAAGTGGCCCCTCTTGACAGCTG 1077
              ||| || || ||||||||||||||||||||||| |||||||| |||||||| || |||||
Db       1223 GAAAATCCCTGGCTTTGTGGCAGCTCAGCCTGTTGACCAAGTCCGGCTCTTAGAAAGCTG 1282

Qy       1078 TTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCT 1137
              |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||||
Db       1283 CTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAGCT 1342

Qy       1138 CATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCT 1197
              ||| || |||| || || |||||||||||||||||||||| ||||||||||| |||||
Db       1343 CATTTTCGCTCCCGACCTCGTTCTGGACAGGCATGAGGGGAAGTGCGTAGAAGGAATTCT 1402

Qy       1198 GGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACA 1257
              |||||||||||||||||||||||| || || ||||||| || ||||||||||||| ||
Db       1403 GGAAATCTTTGACATGCTCCTGGCCACGACGTCAAGGTTCCGTCAGTTAAAACTCCAGCA 1462

Qy       1258 CAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGCT 1317
              ||| || |||||||||| |||||||||||||| ||||| |||||||||||||  |||
Db       1463 CAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTGCC 1522

Qy       1318 CACAGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAAGCCGTGCAC 1377
              | || | ||||| || || || |||||||||||| | ||| | |||||||| |||||
Db       1523 TTCTGCAAACCAGGAGGCAGAAGTAGCCGGAAGCTGACACACCTACTCAAGCCGTGCAC 1582

Qy       1378 CGATGCTTTGGTTTGGGTGATTGCCAAGAGGGGCATCTCCTCCCAGCAGCAATCCATCCG 1437
              |||||  |||| |||||||||| ||||| || ||||||||||||||||| ||  | ||
Db       1583 AGATGCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCG 1642

Qy       1438 CCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGGAGTAACAACAGGGCATGGA 1497
              ||||| |||||||||||||| || || |||||||||||   ||||||||||||||||||
Db       1643 ACTGGCCAACCTCCTGATGCTCCTTTTCCACGTCAGGCACATCAGTAACAACAGGGCATGGA 1702

Qy       1498 ACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGAT 1557
              |||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||
Db       1703 ACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGAT 1762

Qy       1558 GCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCC 1617
              ||||||||| |||  |||||| |||| |||||||| ||| |||||| |||||||| | 
Db       1763 GCTGAATGCTCACAGGCTTCGAGGTACAAGTCCTCAATGTCGGGGTCTGAGTGCAGCTC 1822

Qy       1618 GGCAGAGGACAGTAAAAGCAAACAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGCC 1677
              |||||||||||| | |||||| |||||||||||||||||||||||| | |  |
Db       1823 AACAGAGGACAGTAAGAACAAAGAGACCTCCCAGAACCTACAGTCTCAGTGATGCCAGC 1882

Qy       1678 CCTGAGGTGAACTGGC-CCACAGAGGTCACAAGCTGAAGCGTGAACTCCACTGTCTCAGC 1736
              ||||||| | || | |  || ||| |    ||  || | | || || | ||  ||
Db       1883 CCTGAGGCGGACAGACTACAGAGATGGTCAAAAGTGGAACATGTACCCTAGCATCTGCC 1942


RESULT 28
A61583
LOCUS       A61583                  2568 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 1 from Patent WO9709348.
ACCESSION   A61583
VERSION     A61583.1  GI:3715916
KEYWORDS    .
SOURCE      Rattus rattus.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2568)
  AUTHORS   Kuiper,G.G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348-A 1 13-MAR-1997;
            KAROBIO AB (SE)
FEATURES             Location/Qualifiers
     source          1. .2568
                     /organism="Rattus rattus"
                     /db_xref="taxon:10117"
BASE COUNT      634 a    654 c    663 g    617 t
ORIGIN

  Query Match           55.3%;  Score 1239.6;  DB 5;  Length 2568;
  Best Local Similarity   83.6%;  Pred. No. 9.7e-285;
  Matches 1454;  Conservative    0;  Mismatches  279;  Indels    7;  Gaps    4;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              || || ||| |||||||| ||     | | ||| ||||||||  ||||| |||||||||
Db        213 CATGAGTCTCTGAGAACATAAT--GTCCATCTGTACCTCTTCTCAAAGGAGTTTTCTCA 270

Qy         61 GCTGTTA---TCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTC 117
              |||| |   ||| |||||||||| ||||||||||| || ||||||| ||| ||| |
Db        271 GCTGCGACCCTCTCAAGACATGGAGATCAAAAACTCACCCTCGAGCCTTAGTTC-CCTGC 329

Qy        118 CTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTC 177
              ||||| ||||| || || |||||| |||||||||||||||| |||| |||| |||||
Db        330 TTCCTATAACTGTAGCCAGTCCATCCTACCCCTGGAGCACGGCCCCATCTACATCCCTTC 389

Qy        178 CTCCTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGAT 237
              |||||| ||||||| || |||||| ||| |||| |||||||||| || ||||||||||
Db        390 CTCCTACGTAGACAACCGCCATGAGTATTCAGCTATGACATTCTACAGTCCTGCTGTGAT 449

Qy        238 CAATTACAGCATTCCCAGCAATGTCACTAACTTGCAAGCTGGCCCTGGTCGGCAGACCAC 297
              ||| |||||  |||| ||| |||  || |||| |||||||||||||||||||| |||
Db        450 CAACTACAGTGTTTCCCGCCAGCACCAGTAACCTGGACGGTGGCCCTGTCCGACTCAGCAC 509

Qy        298 AAGCCCAAAIGTGCTATGGCCAACTCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCA 357
              |||||||||||| | ||||||||  |||||||| |||||||| | |||| ||||
Db        510 AAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGCCA 569

Qy        358 GTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGA 417
              | ||| |  || |||||||||||||||||||| ||||||||||||||||| |||||
Db        570 ATCATCCCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGCTGTGAAGCAAGATCACTAGA 629

Qy        418 ACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGC 477
              |||||||||||||||||||||||||||||||||||||||||||||| |||| ||  ||
Db        630 GCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGTGC 689

Qy        478 CAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCCGCTGTCTGCAGCGA 537
              |||||||||||| |||||    ||||||||||||||||||||| | ||||||||||
Db        690 CAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGCGA 749

Qy        538 TTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTAA 597
              ||| |||| || |||| || || || |||| |||||||||||||||||||||||||||
Db        750 TTATGCATCTGGGTATCATTACGGCGTTGGTCATGTGAAGGATGTAAGGCCTTTTTTAA 809

Qy        598 AAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGA 657
              |||||||||||||||||||||||||||||| ||||||| || |||||||||| || ||
Db        810 AAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATAGA 869

Qy        658 TAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAAI 717
              || |||||||| || |||||||||||||||||||||||| |||||||| |||| ||||
Db        870 CAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATCAAGTAGGAAT 929

Qy        718 GGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAG 777
              ||| |||||||| ||| |||||||  | |||||||||||  | |||||||| |||||||
Db        930 GGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGAAG 989

Qy        778 TGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGGCCAGCGCCCCG 837
              | |   ||||||| | |||||      |||| ||||||||||  || || || || |||||
Db        990 ITCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCCCG 1049

Qy        838 AGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGA 897
              |||  |||||| ||||||  | || |||| || |||||| ||||||||||||||||||
Db       1050 GGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTGGA 1109
```

```
Qy        898 GGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCAGTGCGCCCTTCACCGAGGCCTCCAT 957
              ||||| || ||| |||||||| | ||||| |||||    ||||||||||||||||||||
Db       1110 AGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCCAT 1169

Qy        958 GATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGCTACACATGATCAGCTGGCCAA 1017
              |||||||||||| || ||| |||| |||||||  |||| ||||||||  ||||||||||
Db       1170 GATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGCCAA 1229

Qy       1018 GAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTG 1077
              ||| || || |||||||||||||||||||||| |||||||| |||||||| || |||||
Db       1230 GAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGCTG 1289

Qy       1078 TTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCT 1137
              |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||||
Db       1290 CTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAGCT 1349

Qy       1138 CATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCT 1197
              ||| || |||| || || |||||||||||||||||||| |||||||||| |||| |||||
Db       1350 CATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGGATTCT 1409

Qy       1198 GGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACCCCAACA 1257
              |||||||||||||||||||||||| || || |||||||| || ||||||||||||| ||
Db       1410 GGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACCCCAGCA 1469

Qy       1258 CAAAGAATATCTCTGTGTCAAGCCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGT 1317
              ||| || |||||||||||| |||||||||||||| |||| ||||||||||||||  |||
Db       1470 CAAGGAGTATCTCTGTGTGAAGGCCATGATCCTGCTCAACTCCAGTATGTACCCCTTGGC 1529

Qy       1318 CACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGAC 1377
                | || | |||||| || || || |||||||||| | ||| | |||||||| |||||
Db       1530 TTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTGAC 1589

Qy       1378 CGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCG 1437
              |||||  |||| |||||||||||| ||||| || ||||||||||||||||| || | ||
Db       1590 AGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCG 1649

Qy       1438 CCTGGCTAACCCTCCTGATGCTCCTGTCCCACGTCAGCCATGCGAGTAACAAGGGCATGGA 1497
              ||||| |||||||||||||| || || |||||||||||    ||||||||||||||||||
Db       1650 ACTGGCCAACCCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATGGA 1709

Qy       1498 ACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGAT 1557
              |||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||
Db       1710 ACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGAT 1769

Qy       1558 GCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCC 1617
              ||||||||| |||  |||||| |||| |||||||| ||| ||||||| |||||||| |
Db       1770 GCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGCTC 1829

Qy       1618 GGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGC 1677
              |||||||||||| | ||||||| |||||||||||| |||||||||||| | | |
Db       1830 AACAGAGGACAGTAAGAACAAAGAGACGCTCCCAGAACCTACAGTCTCAGTGATGGCCAGG 1889

Qy       1678 CCTGAGGTGAACTGGC CCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGG 1736
              |||||| | || | |  || ||| |   ||  || | | || || | ||  | |  ||
Db       1890 CCTGAGGCGGACAGACTACAGAGATGGTCAAAAGTGGAACATGTACCCTAGCATCTGGGG 1949


RESULT 29
AR075920
LOCUS       AR075920                2568 bp    DNA     linear   PAT 30-AUG 2000
DEFINITION  Sequence 1 from patent US 5958710.
ACCESSION   AR075920
VERSION     AR075920.1  GI:10002666
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2568)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J. A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710 A 1 28 SEP-1999;
FEATURES             Location/Qualifiers
     source          1..2568
                     /organism="unknown"
BASE COUNT      634 a    654 c    664 g    617 t
ORIGIN

  Query Match             65.3%;  Score 1239.6;  DB 6;  Length 2568;
  Best Local Similarity   83.6%;  Pred. No. 9.7e-285;
  Matches 1454;  Conservative    0;  Mismatches  279;  Indels    7;  Gaps    4;

Qy          1 CACGAATCTTTGAGAACATTATAATGACTTTGTGCCTCTTCTTGCAAGGTGTTTCTCA 60
              || || ||| |||||||| ||     | | ||| |||||||  |||| |||||||||
Db        213 CATGAGTCTCTGAGAACATAAT--GTCCATCTGTACCTCTTCTCACAAGGAGTTTTCTCA 270

Qy         61 GCTGTTA---TCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTC 117
              ||||  |   ||| |||||||||| || |||||||||| || ||||||| ||| ||| |
Db        271 GCTGCGACCCTCTGAAGACATGGAGATCAAAAACTCACCGTCGAGCCTTAGTTC-CCTGC 329

Qy        118 CTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTC 177
              ||||| ||||| || || |||||| ||||||||||||||||| |||| |||| |||||
Db        330 TTCCTATAACTGTAGCCAGTCCATCCTACCCCTGGAGCACGGCCCCATCTACATCCCTTC 389

Qy        178 CTCCTATGTAGACAGCCACCATGAATATCCAGCCATCACATTCTATAGCCCTGCTGTGAT 237
              |||||| ||||||| || |||||| ||| |||| |||||||||| || ||||||||||
Db        390 CTCCTACGTAGACAACCGCCATGAGTATTCAGCTATGACATTCTACAGTCCTGCTGTGAT 449

Qy        238 GAATTACAGCATTGCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCAC 297
              ||| |||||  |||| |||    || |||| |||| ||||||||||  || | || |||
Db        450 GAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGCAC 509

Qy        298 AAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCA 357
              |||||||||||| | |||||||  |||||||||| |||||||||| | |||| ||||
Db        510 AAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGCCA 569

Qy        358 GTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGA 417
              | ||| |  || |||| |||||||||||||||| |||||||||||||||| ||||
Db        570 ATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTAGA 629

Qy        418 ACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGC 477
              |||||||||||||||||||||||||||||||||| |||||| |||||||| | |||| ||
Db        630 GCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGTGC 689

Qy        478 CAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGA 537
              |||||||||||| |||||  |||||||||||||||||||||||||| | ||||||||||
Db        690 CAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGCGA 749

Qy        538 TTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAA 597
              ||| |||| || |||| || || || |||| ||||||||||||||||||||||||||||
Db        750 TTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTTAA 809

Qy        598 AAGAAGCATTCAAGGACATAATCATTATATTTGTCCAGCTACAAATCAGTGTACAATCGA 657
              |||||||||||||||||||||||||||||||||| ||||||| || |||||||| || ||
Db        810 AAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATAGA 869

Qy        658 TAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAAT 717
              || |||||||| |||||||||||||||||||||||| |||||||| |||| ||||
Db        870 CAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGAAT 929

Qy        718 GGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAG 777
              ||| |||||||| ||| |||||||  | ||||||||||  | |||||||| |||||||
Db        930 GGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGAAG 989

Qy        778 TGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCG 837
              | |  |||||| | ||||     |||| ||||||||||  || || || || ||||||
Db        990 TTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCCCG 1049

Qy        838 AGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGA 897
              |||  |||||| |||||| | || |||| || |||| || ||||||||||||||||||
Db       1050 GGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTGGA 1109

Qy        898 GGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCAGTGCGCCCTTCACCGAGGCCTCCAT 957
              ||||| || ||| |||||||| | ||||| |||||    ||||||||||||||||||||
Db       1110 AGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCCAT 1169
```

```
Qy        958 CATGATGTCCCTCACCAACTTGGCCGACAAGGAGTTGGTACACATGATCACCTGGGCCAA 1017
              ||||||||||||| || ||| |||| ||||||||  |||| ||||||||  ||||||||||
Db       1170 GATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCCAA 1229

Qy       1018 GAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTG 1077
              ||| || || |||||||||||||||||||||||| |||||||| ||||||| || |||||
Db       1230 CAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGCTG 1289

Qy       1078 TTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCT 1137
              |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||||
Db       1290 CTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAGCT 1349

Qy       1138 CATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCT 1197
              ||| || |||| || || ||||||||||||||||||||| ||||||||||| |||||
Db       1350 CATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGCAAGTGCGTAGAAGGGATTCT 1409

Qy       1198 GGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACA 1257
              |||||||||||||||||||||||| || || |||||||| || ||||||||||||||| ||
Db       1410 GGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAGCA 1469

Qy       1258 CAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGT 1317
              ||| || |||||||||||| |||||||||||||| ||||| ||||||||||||||  |||
Db       1470 CAAGGAGTATCTCTGTGTGAAGGCCATGATCCTGCTCAACTCCAGTATGTACCCCTTGGC 1529

Qy       1318 CACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGAC 1377
              | || | ||||| || || || |||||||||||| | ||| | |||||||| ||||
Db       1530 TTCTGCAAACCAGGAGGCAGAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTGAC 1589

Qy       1378 CGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATCCC 1437
              |||||  |||| |||||||||||| ||||| || ||||||||||||||||| ||  | ||
Db       1590 AGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCG 1649

Qy       1438 CCTGGCTAACCCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGA 1497
              |||||  |||||||||||||| || || ||||||||||    ||||||||||||||||||
Db       1650 ACTGGCCAACCCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATGGA 1709

Qy       1498 ACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGAT 1557
              |||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1710 ACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGAT 1769

Qy       1558 GCTGAATGCCCACGTGCTTCGGGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCC 1617
              ||||||||| |||  |||||| |||| ||||||||| ||| ||||||| |||||||| |
Db       1770 GCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGCTC 1829

Qy       1618 GGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTCACGCCTGGC 1677
              ||||||||||||| | |||||| |||||||||||| ||||||||||||| | |  |
Db       1830 AACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGATGGCCAGG 1889

Qy       1678 CCTGAGGTGAACTGGC-CCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGG 1736
              |||||| | || | |  || ||| |    ||  || | | || || | |  ||
Db       1890 CCTGAGGCGGACAGACTACAGAGATGGTCAAAAGTGGAACATGTACCCTAGCATCTGGGG 1949


RESULT 30
AX022455
LOCUS       AX022455               2568 bp    DNA     linear   PAT 20-SEP-2000
DEFINITION  Sequence 1 from Patent EP0935000.
ACCESSION   AX022455
VERSION     AX022455.1  GI:10046068
KEYWORDS    .
SOURCE      black rat.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae;
            Rattus.
FEATURES             Location/Qualifiers
     source          1. .2568
                     /organism="Rattus rattus"
                     /db_xret="taxon:10117"
BASE COUNT      634 a    654 c    663 g    617 t
ORIGIN

  Query Match             65.3%;  Score 1239.6;  DB 6;  Length 2568;
```

```
  Best Local Similarity   83.6%;  Pred. No. 9.7e-28;
  Matches 1454;  Conservative    0;  Mismatches  279;  Indels    7;  Gaps    4;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              || || ||| |||||||| ||      | | ||| ||||||||  ||||| ||||||||
Db        213 CATGAGTCTCTGAGAACATAAT--GTCCATCTGTACCTCTTCTCACAAGGAGTTTCTCA 270

Qy         61 GCTGTTA---TCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTC 117
              ||||  |   ||| |||||||||| || |||||||||| || ||||||| ||| ||| |
Db        271 GCTGCGAGCCTCTGAAGACATGGAGATCAAAAACTCACCGTCGAGCCTTAGTTC-CCTGC 329

Qy        118 CTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTC 177
              ||||| ||||| || || |||||| ||||||||||||||||| |||| ||||| |||||
Db        330 TTCCTATAACTGTAGCCAGTCCATCCTACCCCTGGAGCACGGCCCCATCTACATCCCTTC 389

Qy        178 CTCCTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGAT 237
              |||||| ||||||| || |||||| ||| |||| |||||||||||| || ||||||||||
Db        390 CTCCTAGGTAGACAACGGCCATGAGTATTCAGCTATGACATTCTACAGTCCTGCTGTGAT 449

Qy        238 GAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCAC 297
              ||| |||||  |||| |||     || |||| |||| ||||||||||  || | || |||
Db        450 GAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGAGGGTGGGCCTGTCCGACTGAGCAC 509

Qy        298 AAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCA 357
              |||||||||||| | ||||||||  |||||||||| |||||||||| | |||| ||||
Db        510 AAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGCCA 569

Qy        358 GTTATCACATCTGTATCCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGA 417
              | ||| |  || |||| |||||||||||||||| |||||||||||||||||| |||||
Db        570 ATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTAGA 629

Qy        418 ACACACCTTACCTGTAAACAGAGAGACACTGAAAGGAAGCTTAGTGGGAACCGTTGCGC 477
              |||||||||||||||||||||||||||||||| |||||| |||||||| | ||| ||
Db        630 GCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGTGC 689

Qy        478 CAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCCCTGTCTGCAGCGA 537
              |||||||||||| |||||    ||||||||||||||||||||||| | ||||||||||
Db        690 CAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGCGA 749

Qy        538 TTACCCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAA 597
              ||| |||| || |||| || || || |||| ||||||||||||||||||||||||||||
Db        750 TTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTTAA 809

Qy        598 AAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGA 657
              |||||||||||||||||||||||||||||||| |||||||| || |||||||||| || ||
Db        810 AAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATAGA 869

Qy        658 TAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAAT 717
              || |||||||| || ||||||||||||||||||||||| |||||||| |||||| |||||
Db        870 CAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGAAT 929

Qy        718 GGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAG 777
              ||| ||||||| ||| ||||||| | ||||||||||  | |||||||| |||||||||
Db        930 GGTCAAGTGTGCATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGAAG 989

Qy        778 TGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAACAGAACTCCGGCCAGCGCCCCG 837
              | |   |||||| | |||||     |||| ||||||||||  || || || || ||||
Db        990 TTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCCG 1049

Qy        838 AGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCGAGCAGCTAGTGCTCACCCTCCTGGA 897
              |||  |||||| |||||  | || |||| || |||| || ||||||||||||||||||
Db       1050 GGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGACCAACTGCTGCTCACCCTCCTGGA 1109

Qy        898 GGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCAT 957
              ||||| || ||| |||||||| | |||| ||||   ||||||||||||||||||||||
Db       1110 AGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCCAT 1169

Qy        958 CATGATGTCCCTCACCAACTTGGCCGACAAGGAGTTGGTACACATGATCACCTGGGCCAA 1017
              |||||||||||| || ||| |||| ||||||||  |||| ||||||||  ||||||||||
Db       1170 GATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCCAA 1229

Qy        1018 GAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTG 1077
               ||| || || |||||||||||||||||||||||| |||||||| ||||||| || |||||
Db        1230 CAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGCTG 1289
```

```
Db        1240 GAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGCTG 1299

Qy        1078 TTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCT 1137
               ||||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||||
Db        1300 CTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAGCT 1359

Qy        1138 CATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCT 1197
               ||| || |||| || || ||||||||||||-||||||||| |||||||||| ||||
Db        1360 CATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGGATTCT 1419

Qy        1198 GGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACA 1257
               |||||||||||||||||||||||| || || |||||||| || ||||||||||||| ||
Db        1420 GGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAGCA 1479

Qy        1258 CAAAGAATATCTCTGTGTCAAGCCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGCT 1317
               ||| || |||||||||| ||||||||||||| ||||| |||||||||||||  |||
Db        1480 CAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTGGC 1539

Qy        1318 CACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGAC 1377
               | || | |||||| || || || ||||||||||| | ||| | |||||||| ||||
Db        1540 TTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTGAC 1599

Qy        1378 CGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGGG 1437
               |||||  |||| |||||||||||| ||||| || |||||||||||||||||| ||  | ||
Db        1600 AGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCG 1659

Qy        1438 CCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGA 1497
               ||||| |||||||||||| || || ||||||||||    |||||||||||||||||||
Db        1660 ACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATGGA 1719

Qy        1498 ACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCACTGTATGACCTGCTGCTGGAGAT 1557
               |||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||
Db        1720 ACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGAT 1779

Qy        1558 GCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCC 1617
               ||||||||| ||| |||||| |||| |||||||||| ||| ||||||| |||||||||
Db        1780 GCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTCAGTGCAGCTC 1839

Qy        1618 GGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGC 1677
               |||||||||||| | ||||||| |||||||||| |||||||||||| | |  |
Db        1840 AACAGAGGACAGTAAGAACAAAGAGACGCTCCCACAACCTACAGTCTCAGTGATGGCCAGG 1899

Qy        1678 CCTGAGGTGAACTGGC CCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGG 1736
               ||||||| | || | |  || ||| |    ||  || | | || || | ||  | |  ||
Db        1900 CCTGAGGCGGACAGACTACAGAGATGGTCAAAAGTGGAACATGTACCCTAGCATCTGGGG 1959
```

```
RESULT 41
AF164957
LOCUS       AF164957               1581 bp    mRNA    linear   MAM 19-APR-2002
DEFINITION  Sus scrofa estroqen receptor beta mRNA, complete cds.
ACCESSION   AF164957
VERSION     AF164957.2  GI:14717426
KEYWORDS    .
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G., Vale-Cruz,D.S., Choi,I.,
            Katzenellenbogen,B.S., Simmen,F.A. and Simmen,R.C.
  TITLE     Molecular cloninq of porcine estroqen receptor-beta complementary
            DNAs and developmental expression in perilmplantation embryos
  JOURNAL   Biol. Reprod. 66 (4), 760-769 (2002)
  MEDLINE   21858433
   PUBMED   11870084
REFERENCE   2  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G. and Simmen,R.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1999) Animal Science, University of Florida,
            Shealy Drive, Buildinq 459, Gainesville, FL 32611, USA
REFERENCE   3  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G. and Simmen,R.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2001) Animal Science, University of Florida,
            Shealy Drive, Buildinq 459, Gainesville, FL 32611, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 13, 2001 this sequence version replaced qi:5566377.
FEATURES             Location/Qualifiers
     source          1. .1581
                     /orqanism "Sus scrofa"
                     /db_xref="taxon:9823"
                     /dev_staqe="embryo"
     CDS             1. .1581
                     /codon_start 1
                     /product="estroqen receptor beta"
                     /protein_id "AAD45381.2"
                     /db_xref="GI:14717427"
                     /translation="MDIKNSPSNLNSPVSYNCSQSVLPLEPGPTYIPSSYVESCHEYS
                     AMTFYSPAVVNYSISSNSEVGPGRQATSPNVLWPTPGHLSPLATHCQPSLLYAEPQKS
                     PWCDTRSLEHTLPVNRETLKRKASGSSCASPVTSPSSKRDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRR
                     ERCGYRIVRKQRNSEGHLHCLSRAKKNGDHTTRVKELLLSTLSPEQLVLTLLEAEPPH
                     VLVSRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFMELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHK
                     EYLCVKAMILLNSSMYPSAAAQEAESSRKLTHLLNAVIDALVWVIARSGISSQQQSVR
                     LANLLMLLSHVRHASNKGTEHLLNMKCKNVVPVYDLLLEMLNAHTLRGNKSLVTGSER
                     SRMEESESKEGSQKPQAQ"
BASE COUNT      496 a     424 c     411 q     450 t
ORIGIN
```

```
  Query Match             64.7%;  Score 1227.4;  DB 4;  Length 1581;
  Best Local Similarity   86.6%;  Pred. No. 7.5e-282;
  Matches 1380;  Conservative    0;  Mismatches  201;  Indels   12;  Gaps    2;

Qy          77 ATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAA 136
               ||||||||||||||||||||||||||||||| |||||| ||||||  ||||||||||||||||||
Db           1 ATGGATATAAAAAACTCACCATCTAACCTTAACTCTCCTGTCTCCTACAACTGCAGTCAA 60

Qy         137 TCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCAC 196
               ||  || | ||||||||||  ||| |||||||| || |||||||||||||||||| |||  |
Db          61 TCTGTCCTGCCCCTGGAGCCAGGCCCCATATATATCCCTTCCTCCTATGTAGAAAGCTGC 120

Qy         197 CATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGCAATTACAGCATTCCCAGC 256
               ||||||||| |||||||||||| ||||||||||||||| |||||||||| ||| ||||
Db         121 CATGAATATTCAGCCATGACATTTTATAGCCCTGCTGTGGTGAATTACAGTATTTCCAG- 179

Qy         257 AATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGG 316
                     |||||  |||| |||||||||||| ||| |||||||||| ||||||||||
Db         180  - -----TAACTCCGAAGTTGGGCCTGGTCGACAGGCCACAAGCCCCAATGTGTTGTGG 231

Qy         317 CCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCG 376
               ||||| ||||| ||||| |||||||||   ||||| |||||   |||| |||||||||
Db         232 CCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCACTTCTGTATGCA 291

Qy         377 GAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAAC 436
               |||||||||||||||| |||||||  ||||||| |||||||||||||||||||||| |||
Db         292 GAACCTCAAAAGAGTCCTTGGTGTGACACAAGATCACTAGAACACACCTTACCTGTGAAC 351

Qy         437 AGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCA 496
               ||||||||| |||||||||||| |||||||| | |||| |||||||||||||  |||||
Db         352 AGAGAGACATTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGTTACCAGTCCA 411

Qy         497 GGTTCAAACAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCAC 556
               |||||||||||||||| |||||||| |||||||||| ||||| || || |||||||||
Db         412 AGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCAGGATATCAC 471

Qy         557 TATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACAT 616
               |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db         472 TATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACAT 531

Qy         617 AATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGC 676
               |||||||||||||||||||||||||||| |||||||| || ||||| |||||||| || |||
Db         532 AATGATTATATTTGTCCAGCTACAAACCAGTGTACGATAGATAAGAACCGGCGTAAAAGC 591

Qy         677 TGCCAGCGCCTGCGCACTTCGGAAGTGTTACGAAGTGGGAATGCTGAAGTGTGGCTCCGG 736
```

```
Db        592 TGCCAGGCCTGCCGACTTCGAAAGTGCTATGAAGTCGGAATGGTGAAGTGTGGCTGCCGC 651

Qy        737 AGAGAGAGATGTGGGTACCGCTTGTGCGGAGACAGACAAGTGCCCACGAGCAGCTGCAC 796
              |||||  | ||||||||||| | |||| |  |||||| | | || | ||| ||||||
Db        652 AGAGAACGGTGTGGGTACCGCATAGTGCGAAAGCAGAGAAATTCTGAGGGGCATCTGCAC 711

Qy        797 TGTGCCGCAAGGCCAAGAGAAGTGCGCGCAGCGCCCGAGTGCGGAGCTGCTGCTG 856
              ||    |||  |||||||  | ||| |  ||| |  ||||||||  |||||||| |||
Db        712 TGCCTGAGCAGAGCCAACAAGAATGGTGATCACACAACCCCAGTCAAGGAGCTGCTGCTG 771

Qy        857 GACGCCCTGAGCCCCGAGCAGCTAGTGCTCACGCTCCTGGAGGCTGAGCCGCCCCATCCC 916
              | ||||||||||  |||||||| ||||||| ||||||||||  |||| |||||  |||
Db        772 AGCACCCTGAGCCCAGAGCAGCTGGTGCTCACGCTCCTGGAGGCCGAGCCACCCCACGTG 831

Qy        917 CTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAG 976
              ||| | ||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db        832 CTGGTGAGCCGCCCCAGCAGCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAG 891

Qy        977 TTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTG 1036
              ||||||||||||||  |||||||||||||||||||||||||||| |||||  ||||  ||
Db        892 CTGCCCGACAAGCAACTGGTACACATGATCAGCTGGGCCAAGAAAATTCCGGGCTTCACC 951

Qy       1037 GAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGACGTGTTGGATGGAGGTGTTAAT 1096
              ||||||||||||| ||||||||||||||| |||||||||  ||| |||||||| | |||
Db        952 GAGCTCAGCCTGTACGACCAAGTGCGGCTCCTGGAGACGTGCTGGTTGCAGGTGCTCAT 1011

Qy       1097 ATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTT 1156
              |||||||||||||||||| || ||||||||||||||||||||||||||||||||| ||
Db       1012 GTGGGGCTGATGTGGCGCTGCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTC 1071

Qy       1157 GTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGCAAATCTTTGACATGCTC 1216
              |||||||||||||||| |||||||| ||||||||||||||||||||||||||||||||
Db       1072 GTTCTGGACAGGGATGAAGGGAAATGTGTAGAAGGAATTCTGCAAATCTTTGACATGCTC 1131

Qy       1217 CTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTC 1276
              ||||||| |||||||||||||| ||||||||||||||||||| || ||||||||||||
Db       1132 CTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAGGAGTATCTCTGTGTC 1191

Qy       1277 AAGGCCATGATCCTGCTCAATTCCAGTATGTACCTCTGGTCACAGGGAGCCAGGATGTT 1336
              |||||||||||||| |||| |||||||||| |||   |  |||   |||| || |||
Db       1192 AAGGCCATGATCCTTCTCAACTCCAGTATGTATCTTCAGCTCCAG----CCCAACAGGAT 1248

Qy       1337 GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTGGGGTG 1396
              || |||||||||||||| ||||  |||||| |||||||| |||||  |||| ||||||
Db       1249 GAGAGCAGCCGGAAGCTGACTCATCTGCTGAATGCCGTGACTGATGCCTTGGTCTGCGTC 1308

Qy       1397 ATTGCCAACAGCGGCATGTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATG 1456
              || |||| ||| || |||||||||||||||| ||  | |||||||| |||||||| |||
Db       1309 ATCGCCAGGAGTGGTATGTCCTCCCAGCAGCAGTCTGTGCGCCTGGCCAACCTCCTCATG 1368

Qy       1457 CTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAG 1516
              ||||| || || ||||| ||||| |||||||||||||| ||||||  ||||||||||||||
Db       1369 CTCCTTTCTCATGTCAGACATGCCAGTAACAAGGGCACGGAACACTTGCTCAACATGAAG 1428

Qy       1517 TGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCCT 1576
              |||||||||||||||| || || ||||||||||||||||||||||||||||||  |||
Db       1429 TGCAAAAATGTGGTCCCCGTCTACGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCCT 1488

Qy       1577 CGCGGGTGCAAGTCCTCCATCGAGGAGTCCGAGTCAGCCCCGGCAGAGGACAGTAAAAGC 1636
              || ||   |||||||  |||||||||| ||| |||||| |   |||| ||| | |||
Db       1489 CGTGGCAACAAGTCCTTGGTCACGGGGTCTGAGGCGAGCCGGATGGAGGAGAGTGAGAGC 1548

Qy       1637 AAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
              |||||||||||||||| || ||| ||||||||
Db       1549 AAAGAGGGCTCCCAGAAACCCAGGCTCAGTGA 1581


RESULT 32
AF267736
LOCUS       AF267736               1581 bp    mRNA    linear   MAM 15-MAR-2002
DEFINITION  Sus scrofa estrogen receptor beta (ER beta) mRNA, complete cds.
ACCESSION   AF267736
VERSION     AF267736.1  GI:13183320
KEYWORDS    .
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1581)
  AUTHORS   LaVoie,H.A., Desimone,D.C., Gillio-Meina,C. and Hui,Y.Y.
  TITLE     Cloning and characterization of porcine ovarian estrogen receptor
            beta isoforms
  JOURNAL   Biol. Reprod. 66 (3), 616-623 (2002)
  MEDLINE   21858315
   PUBMED   11870066
REFERENCE   2  (bases 1 to 1581)
  AUTHORS   LaVoie,H.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2000) Cell Biology and Neuroscience, University
            of South Carolina, School of Medicine, Bldg 1, Columbia, SC 29208,
            USA
FEATURES             Location/Qualifiers
     source          1. .1581
                     /organism="Sus scrofa"
                     /db_xref "taxon:9823"
                     /tissue_type "ovary"
     gene            1. .1581
                     /gene="ER beta"
     CDS             1. .1581
                     /gene="ER beta"
                     /codon_start 1
                     /product="estrogen receptor beta"
                     /protein_id="AAK15151.1"
                     /db_xref "GI:13183321"
                     /translation="MDIKNSPSNLNSPVSYNCSQSVLPLEPGPIYIPSSYVESCHEYS
                     AMTFYSPAVVNYSISSNSFVGPGPQATSPNVLWPTPGRLSPLAIHCQPSIYAEPQKS
                     PWCDTRSLEHTLPVNRETLKRKASGSSCASPVTSPSSKRDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGSPR
                     ERCGYRIVPKQPNSEGHLHCLSPAKKNGDHTTPVKELLSTLSPEQLVLTLLEAEPPH
                     VLVSPPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHK
                     EYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSVR
                     LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGNKSTVTGSER
                     SEMEESSSKEGSQKPQAQ"
BASE COUNT      394 a    425 c    412 g    350 t
ORIGIN

  Query Match             64.7%;  Score 1227.4;  DB 4;  Length 1581;
  Best Local Similarity   86.6%;  Pred. No. 7.5e-282;
  Matches 1380,  Conservative    0,  Mismatches  201,  Indels   12,  Gaps    2;

Qy          77 ATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAA 136
               ||||||| |||||||||||||||| |||||| ||||||  |||||||||||||||||||
Db           1 ATGGATATCAAAAACTCACCATCTAACCTTAACTCTCCTGCTCCTACAACTGCAGTCAA 60

Qy         137 TCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCAC 196
               ||  || | |||||||||  ||| |||||| || ||||||||||||||||| ||| | |
Db          61 TCTGTCCTGCCCCTGGAGCCAGGCCCCATATATATCCCTTCCTCCTATGTACAAAGCTGC 120

Qy         197 CATGAAATATCCAGCCATGACATTCTATAGCCCTGCTGTGCATGAATTACAGCATTCCCAGC 256
               ||||||||| |||||||||||| ||||||||||||| |||||||||| ||| ||||
Db         121 CATGAAIATTCAGCCATGACATTTIATAGCCCTGCTGTGGTGAATTACAGIATTCCCAG- 179

Qy         257 AATGTCACTAACCTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGG 316
                     |||||  |||| |||||||||||| ||| ||||||||||  ||||||||||||
Db         180 --------TAACTCCGAAGTTGGGCCTGGTCGACAGGCCACAAGCCCCAATGTGTTGTGG 231

Qy         317 CCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCG 376
               ||||| |||| |||| ||||||||||   ||||| |||||   |||| |||||||||
Db         232 CCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCACTTCTGTATGCA 291

Qy         377 GAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATGCTAGAACACACCTTACCTGTAAAC 436
               |||||||||||||||| |||||||  |||||| |||||||||||||||||||||| |||
Db         292 GAACCTCAAAAGAGTCCTTGGTGTGACACAAGATCACTAGAACACACCTTACCTGTGAAC 351
```

```
Qy        437 AGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCA 496
              ||||||||| |||||||||||| |||||||| | |||| ||||||||||||||  |||||
Db        452 AGAGAGACATTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGTTACCAGTCCA 511

Qy        497 GGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCAC 556
              |||||||||||||||| |||||||| |||||||||| ||||| || || |||||||||
Db        512 AGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCAGGATATCAC 571

Qy        557 TATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACAT 616
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        572 TATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACAT 631

Qy        617 AATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGC 676
              |||||||||||||||||||||||||| |||||||| || |||| |||||||| || |||
Db        632 AATGATTATATTTCTCCAGCTACAAACCAGTGTACGATAGATAAGAACCGGCGTAAAAGC 691

Qy        677 TGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGG 736
              |||||||||||||||||||| |||| || |||| ||||||||||||||||||||||||
Db        692 TGCCAGGCCTGCCGACTTCGAAAGTGCTATGAAGTCGGAATGGTGAAGTGTGGCTCCCGG 751

Qy        737 AGAGAGAGATGTGGGTACCGCCTTGTGCCGGAGACAGAGAAGTGCCGACGAGCAGCTGCAC 796
              |||||  | |||||||||||| | |||| |  |||||||  | || | ||| |||||||
Db        752 ACAGAACGGTGTGGGTACCGCATAGTGCGAAAGCAGAGAAATTCTGAGGGGCATCTGCAC 811

Qy        797 TGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTG 856
              ||    ||| |||||||  | ||| |  ||| |  |||||||  |||||||| |||
Db        812 TGCCTGAGCAGAGCCAAGAAGAATGGTGATCACACAACCCGAGTGAAGGAGCTGCTCCTG 771

Qy        857 GACGCCCTGAGCCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTG 916
              | |||||||||| |||||||| |||||||| |||||||||| ||||| ||||| |||
Db        772 ACCACCCTCAGCCCACAGCAGCTGGTGCTCACGCTCCTGGAGGCCGAGCCACCCCAGGTG 831

Qy        917 CTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAG 976
              ||| | ||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db        832 CTGGTGAGCCGCCCCAGCACGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAG 891

Qy        977 TTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTG 1036
              |||||||||||||  |||||||||||||||||||||||||||| ||||| ||||| |||
Db        892 CTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAAATTCCGGGCTTCGTG 951

Qy        1037 GAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATG 1096
               |||||||||||| |||||||||||||| |||||||||| ||| |||||||| | |||
Db        952 GAGCTCAGCCTGTACGACCAAGTGCGGCTCCTGGAGAGCTGCTGGTTGGAGGTGCTCATG 1011

Qy        1097 ATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTT 1156
               |||||||||||||||||| || |||||||| ||||||||||||||||||||||||||| ||
Db        1012 GTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTC 1071

Qy        1157 GTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTC 1216
               |||||||||||||||| |||||||| ||||||||||||||||||||||||||||||||
Db        1072 GTTCTGGACAGGGATGAAGGGAAATGTGTAGAAGGAATTCTGGAAATCTTTGACATGCTC 1131

Qy        1217 CTGGCAACTACTTCAACGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTC 1276
               ||||||| |||||||||||||| |||||||||||||||||| ||||||||||||||||
Db        1132 CTGGCAACGACTTCAACGTTTCGTGAGTTAAAACTCCAACACAAGGAGTATCTCTGTGTC 1191

Qy        1277 AAGGCCATGATCCTGCTCAACTCCAGTATGTATCCTTCAGCTGCAG---CCCAAGAGGCT 1333
               |||||||||||||| ||||| |||||||||| |||   |  |||   |||| || |||
Db        1192 AAGGCCATGATCCTCCTCAACTCCAGTATGTATCCTTCAGCTGCAGAAACCCAAGAGGCT 1248

Qy        1334 GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTGGGTG 1393
               || |||||||||||||| |||| |||||||| |||||||| |||||  |||| ||||||
Db        1249 GAGAGCAGCCGCAAGCTGACTCATCTGCTGAATGCCGTGACTGATGCCCTGGTCTGGGTG 1308

Qy        1394 ATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATG 1453
               || |||| ||| || |||||||||||||||||| ||  | |||||||| |||||||| |||
Db        1309 ATCGCCAGGAGTGGTATCTCCTCCCAGCAGCAGTCTGTCCGCCTGGCCAACCTCCTCATG 1368

Qy        1454 CTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAG 1513
               ||||| || || |||| |||| ||||||||||||||||||  ||||||||||||||||
Db        1369 CTCCTTTCTCATGTCAGACATGCCAGTAACAAGGGCATGGAACACTTGCTCAACATGAAG 1428

Qy        1514 TGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTT 1573
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1429 TGCAAAAATGTGGTCCCCGTCTACGACCTGCTGCTGGAGATGCTGAATGCCCACACTCTT 1488

Qy        1574 CGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGC 1633
               || ||   |||||||| ||||||||| || |||||  |  |||| ||| | |||
Db        1489 CGTGGCAACAAGTCCTTGGTCACGGGGTCTGAGCGCAGCCGGATGGAGGAGAGTGAGAGC 1548

Qy        1634 AAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
               |||||||||||||||| || ||| ||||||||
Db        1549 AAAGAGGGCTCCCAGAAACCCCAGGCTCAGTGA 1581


RESULT 33
BC024181
LOCUS       BC024181                2153 bp    mRNA    linear   PRI 21-FEB-2002
DEFINITION  Homo sapiens, clone MGC:26496 IMAGE:4825947, mRNA, complete cds.
ACCESSION   BC024181
VERSION     BC024181.1  GI:18848207
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2153)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 33 Row: m Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 10845012.
FEATURES             Location/Qualifiers
     source          1. .2153
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "MGC:26496 IMAGE:4825947"
                     /tissue_type="Testis"
                     /clone_lib "NIH_MGC_97"
                     /lab_host "DH10B"
                     /note "Vector: pBluescript"
     CDS             291. .1262
                     /codon_start 1
                     /product="Unknown (protein for MGC:26496)"
                     /protein_id "AAH24181.1"
                     /db_xref "GI:18848208"
                     /translation "MDTKNSPSSLNSPSSYNCSQSTLPTLEHGSTYTPSSYVDSHHEYP
                     AMTFYSPAVMNYSTPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEARSLEHTLPVNREILKRKVSGNPCASPVTGPGSNKRDAHECAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNKRKSCQACRLRKCYEVGMVKCG
                     SRRERGGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGMRGNA"
BASE COUNT      563 a    528 c    549 g    513 t
ORIGIN

  Query Match             62.9%;  Score 1194;  DB 9;  Length 2153;
  Best Local Similarity   90.6%;  Pred. No. 7.2e-274;
```

```
  Matches 1343;  Conservative    0;  Mismatches    0;  Indels  139;  Gaps    1;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 274

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 334

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 394

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 454

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 514

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 574

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 634

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 694

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 754

Qy        541 CGCATCGGGATATCACTATGCAGTCTGGTCGTCTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 CGCATCGGGATATCACTATGCAGTCTGGTCGTCTGAAGGATGTAAGGCCTTTTTTAAAAG 814

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        815 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 874

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCCACTTCGGAAGTGTTACGAAGTGCGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        875 AAACCGGCGCAAGAGCTGCCAGGCCTGCCCACTTCGGAAGTGTTACGAAGTGCGAATGGT 934

Qy        721 GAAGTGTGGCTCCCGCAGAGAGAGATGTGGGTACCGCCTTGTGCGCACACACACAACGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        935 GAAGTGTGGCTCCCGCAGAGAGAGATGTGGGTACCGCCTTGTGCGCACACAGAGAACTGC 994

Qy        781 CGACGAGCAGCTGCACTGTGCCGCCAAGGCCAAGAGAAGTGCGCCGCCACGCGCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        995 CGACGAGCAGCTGCACTGTGCCGCCAAGGCCAAGAGAAGTGCGCCGCCACGCGCCCGAGT 1054

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1055 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 1114

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCGCGCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1115 TGAGCCGCCCCATGTGCTGATCAGCGCGCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 1174

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1175 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCCAAGAA 1234

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGCAGAGCTGTTG 1080
              |||||||
Db       1235 GATTCCC----------------------------------------------------- 1241

Qy       1081 GATGGAGGTGTTAATGATGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 ----------------------------------------------------------- 1241

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
                                                 ||||||||||||||||||||||||||||
Db       1242 -------------------------GGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1275

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1335

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1336 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1395

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1455

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1456 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1515

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1516 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1557


RESULT 34
AF124790
LOCUS       AF124790               1427 bp    mRNA    linear   PRI 27-MAY-1999
DEFINITION  Homo sapiens estrogen receptor beta2 splice variant mRNA, complete
            cds.
ACCESSION   AF124790
VERSION     AF124790.1  GI:4894662
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1427)
  AUTHORS   Lu,B., Leyque,E., Dotzlaw,H., Watson,P.H. and Murphy,L.C.
  TITLE     Cloning and functional characteristics of human estrogen receptor
            beta isoforms
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1427)
  AUTHORS   Lu,B., Leyque,E., Dotzlaw,H., Watson,P.H. and Murphy,L.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Biochemistry, University of Manitoba, 770
            Bannatyne Avenue, Winnipeg, MB R3E 0W3, Canada
FEATURES             Location/Qualifiers
     source          1. .1427
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="ovary"
     CDS             52. .1023
                     /note="missing exon 5"
                     /codon_start=1
                     /product "estrogen receptor beta2 splice variant"
                     /protein_id="AAD32580.1"
                     /db_xref="GI:4894663"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEAKSLEHTLPVNRETLKRKVSGNRGASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SRRFRGGYKQLVKRQRSADEGLHCAGKAKKSGGHAPEVRPLTTDAISPEQLVLTLLEAE
                     PPHVLISPPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFRGNA"
BASE COUNT      457 a     384 c     364 g     322 t
ORIGIN

  Query Match             61.6%;  Score 1169;  DB 9;  Length 1427;
```

```
  Best Local Similarity   90.5%;  Pred. No. 6.4e-268;
  Matches 1318;  Conservative    0;  Mismatches    0;  Indels  139;  Gaps    1;

Qy         26 GACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATA 85
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATA 60

Qy         86 AAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTA 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTA 120

Qy        146 CCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATAT 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATAT 180

Qy        206 CCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACCT 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACCT 240

Qy        266 AACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCT 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCT 300

Qy        326 GGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAA 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAA 360

Qy        386 AAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACA 420

Qy        446 CTCAAAAGGAAGGTTAGTGGGAACCGTTGCGCCACGCCCTGTTACTGGTCCAGGTTCAAAG 505
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCAAAAGGAAGGTTAGTGGGAACCGTTGCGCCACGCCCTGTTACTGGTCCAGGTTCAAAC 480

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 540

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGTCGTGTGAACGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 600

Qy        626 ATTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 660

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 720

Qy        746 TGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGC 805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGC 780

Qy        806 AAGGCCAAGAGAAGTGCCGCCACGCGCCCCGAGTCCGGGAGCTGCTGCTGGACGCCCTG 865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGGCCAAGAGAAGTGCCGCCACGCGCCCCGAGTCCGGGAGCTGCTGCTGGACGCCCTG 840

Qy        866 AGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGC 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGC 900

Qy        926 CGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGAC 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGAC 960

Qy        986 AAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGC 1045
              ||||||||||||||||||||||||||||||||||||| |||||
Db        961 AAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC     - --- - - -  1002

Qy       1046 CTGTTCGACCAAGTGCTGCTCTTGGAGATGTGTTGGATGGAGGTGTTAATGATGGGGCTG 1105
```

```
Db       1003 ------ ----- - - -  ----- -  - - - - - - - - - -  1002

Qy       1106 ATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGAC 1165

Db       1003 ---------- - ---------  ---- ----- - -  -  1002

Qy       1166 AGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACT 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003  GGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACT 1061

Qy       1226 ACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATG 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 ACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATG 1121

Qy       1286 ATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGC 1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 ATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGC 1181

Qy       1346 CGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAG 1405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 CGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAG 1241

Qy       1406 AGCGGCATCTCCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCC 1465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 AGCGGCATCTCCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCC 1301

Qy       1466 CACGTCAGGCATGCGAG 1482
              |||||||||||||||||
Db       1302 CACGTCAGGCATGCGAG 1318
```

```
RESULT 45
AF042058
LOCUS       AF042058                1621 bp    mRNA    linear   ROD 22-JAN-1998
DEFINITION  Rattus norvegicus estrogen receptor beta1 (ERbeta) mRNA, complete
            cds.
ACCESSION   AF042058
VERSION     AF042058.1  GI:2801690
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1621)
  AUTHORS   Petersen,D.N., Tkalcevic,G.T., Koza Taylor,P.H., Turi,T.G. and
            Brown,T.A.
  TITLE     Identification of estrogen receptor beta2, a functional variant of
            estrogen receptor beta expressed in normal rat tissues
  JOURNAL   Endocrinology (1998) In press
REFERENCE   2  (bases 1 to 1621)
  AUTHORS   Brown,T.A. and Petersen,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1998) Molecular Sciences, Pfizer Central
            Research, Eastern Point Road, Groton, CT 06340, USA
FEATURES             Location/Qualifiers
     source          1. .1621
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1621
                     /gene "ERbeta"
     CDS             1. .1458
                     /gene="ERbeta"
                     /note="steroid hormone receptor"
                     /codon_start 1
                     /product="estrogen receptor beta1"
                     /protein_id "AAB97424.1"
                     /db_xref="GI:2801691"
                     /translation="MTFYSPAVMNYSVPGSTSNLIGGPVRQSTSPNVLWPTSGHLSPL
                     ATHCQSSLLYAEPQKSPWCEARSLEHTLPVNRETLKRKLSGSSCASPVTSPNAKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LRKCYEVGMVKCGSRRERCGYRIVRRQRSSSEQVHCLSKAKRNGGHAPRVKELLLSTL
```

```
                     SPEQIVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDREEGKCVEGILEIFDM
                     LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLTHLLNAVTDAL
                     VWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHTLRGYKSSISGSECSSTEDSKNKESSQNLQSQ"
     variation       80
                     /gene="ERbeta"
                     /note="encodes leucine residue in GenBank Accession Number
                     U57439"
                     /replace="t"
     variation       313
                     /gene "ERbeta"
                     /note="encodes proline residue in GenBank Accession Number
                     U57439"
                     /replace="c"
BASE COUNT      404 a    417 c    449 g    351 t
ORIGIN

  Query Match             59.5%;  Score 1130;  DB 10;  Length 1621;
  Best Local Similarity   84.2%;  Pred. No. 1.3e-258;
  Matches 1285;  Conservative    0;  Mismatches  240;  Indels    1;  Gaps    1;

Qy        212 ATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCAGTAACTTG 271
              ||||||||||| || |||||||||||||| |||||  |||||  |||     || |||| ||
Db          1 ATGACATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTG  60

Qy        272 GAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCAC  331
              || |||||||||||  || |||| ||||||||||||||||| | |||||||  |||||||
Db         61 GACGGTGGGCCTGTCCGACAGAGCACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCAC  120

Qy        332 CTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGT  391
              || |||||||||| |  |||| |||| | ||| |   || |||| ||||||||||||||||
Db        121 CTGTCTCCTTTAGCGACCCATTGCCAATCATCGCTCCTCTATGCAGAACCTCAAAAGAGT  180

Qy        392 CCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAA  451
              || ||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||
Db        181 CCTTGGTGTGAAGCAAGATCACTAGAGCACACCTTACCTGTAAACAGAGAGACACTCAAC  240

Qy        452 AGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGAT  511
              |||||| |||||||| | |||| ||||||||||||||| |||||    ||||||||||||
Db        241 AGGAAGCTTAGTGGGAGCAGTTGTGCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGAT  300

Qy        512 GCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCG  571
              |||||||||||||| ||||||||||||||| ||||| || |||| || || || |||||
Db        301 GCTCACTTCTGCGCCGTCTGCAGCGATTATGCATCTGGGTATCATTACGGCGTTTGGTCA  360

Qy        572 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGT  631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        361 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATCTGT  420

Qy        632 CCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGA  691
              ||||| || |||||||||| || || || |||||||| || ||||||||||||||||||||
Db        421 CCAGCCACGAATCAGTGTACCATAGACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGA  480

Qy        692 CTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGCC  751
              ||||| |||||||| |||| |||||||| |||||||| ||| |||||||  | ||||||
Db        481 CTTCGCAAGTGTTATGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGACAACCGTGTGCC  540

Qy        752 TACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCC  811
              ||||| | |||||||| |||||||| |   ||||||| | ||||      |||| |||
Db        541 TACCGTATAGTGCCGGAGGCAGAGAAGTTCTACCGAGCAGGTACACTGCCTGAGCAAAGCC  600

Qy        812 AAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACCCCCTCAGCCCC  871
              |||||||  || || || || |||| |||  |||||| ||||||  | || |||| ||
Db        601 AACAGAAACGGTGGGCATCCACCCCGGGTGAAGCAGCTACTGCTGAGCACCTTGAGTCCA  660

Qy        872 GAGCAGCTAGTGCTCACCCTCCTGGAGGCTCAGCCCCCCCATGTGCTGATCAGCCCCCCC  931
              ||||| || |||||||||||||||||| ||||| || ||| |||||||| | ||||| |||
Db        661 GAGCAACTGGTGCTCACCCTCCTGGAAGCTGAACCACCCAATGTGCTGCTGAGCCGCCCC  720

Qy        932 AGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAG  991
              ||   |||||||||||||||||||||||||||||||| || ||| |||| ||||||||
Db        721 AGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAA  780

Qy        992 TTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTC 1051
              |||| ||||||||  |||||||||||||| || || ||||||||||||||||||||||||
Db        781 CTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTC  840

Qy       1052 GACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGG 1111
              |||||||| |||||||| || |||| |||||||||||| |||| |||| |||||||||
Db        841 GACCAAGTCCGGCTCTTAGAAAGCTGCTGGATGGAGGTGTAATGCTGGGACTGATGTGG  900

Qy       1112 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGAT 1171
              ||||| || ||||||||||||||||||| || |||| || || ||||||||||||||||
Db        901 CGCTCCATCGACCACCCCGGCAAGCTCATTTTCGCTCCCGACCTCGTTCTGGACAGGGAT  960

Qy       1172 GAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCA 1231
              ||||||| |||||||||| ||||||||||||||||||||||||||||||||| || || |||
Db        961 GAGGGGAACTGCGTAGAAGGATTCTGGAAATCTTTGACATGCTCCTGGCGACGACGTCA 1020

Qy       1232 AGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1291
              ||||| || |||||||||||||| ||||| || |||||||||| ||||||||||||||
Db       1021 AGGTTCCGTCAGTTAAAACTCCAGCACAAGGAGTATCTCTGTGTCAAGGCCATGATCCTC 1080

Qy       1292 CTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAG 1351
              ||||| |||||||||||||| |||   | || | |||||| || || || ||||||||
Db       1081 CTCAACTCCAGTATGTACCCCTTGGCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAG 1140

Qy       1352 CTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGC 1411
              ||| | ||| | |||||||| |||| |||||  |||| ||||||||||| ||||| ||
Db       1141 CTGACACACCTACTGAACGCGGTGACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGT 1200

Qy       1412 ATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTC 1471
              |||||||||||||||| || | || |||| |||||||||||||| || || |||||
Db       1201 ATCTCCTCCCAGCAGCAGTCAGTCCGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTC 1260

Qy       1472 AGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTC 1531
              |||||    |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1261 AGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTC 1320

Qy       1532 CCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCC 1591
              || |||||||||||||||||||||||||||||||| |||  |||||| |||| ||||||
Db       1321 CCGGTGTATGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCC 1380

Qy       1592 TCCATCACGGGGTCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAG 1651
              || ||| |||||| |||||||| |  |||||||||||| | |||||| ||||||||
Db       1381 TCAATCTCGGGGTCTGAGTGCAGCTCAACAGAGGACAGTAAGAACAAACAGAGCTCCCAG 1440

Qy       1652 AACCCACAGTCTCAGTGACGCCTGGCCCTGAGGTGAACTGGC-CCACAGAGGTCACAAGC 1710
              |||| |||||||||||| | |  | |||||| | | | | |  || ||| |    ||
Db       1441 AACCTACAGTCTCAGTGATGGCCAGGCCTGAGGCGGACAGACTACAGAGATGGTCAAAAG 1500

Qy       1711 TGAAGCGTGAACTCCAGTGTGTCAGG 1736
              || | | || || | ||  | |  ||
Db       1501 TGGAACATGTACCCTAGCATCTGGGG 1526


RESULT 36
MMU81451
LOCUS       MMU81451                2152 bp    mRNA    linear   ROD 01-APR-1997
DEFINITION  Mus musculus estrogen receptor beta (Estrb) mRNA, complete cds.
ACCESSION   U81451
VERSION     U81451.1  GI:1912467
KEYWORDS    nuclear receptor; steroid receptor; transcription factor; mouse
            chromosome 12.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2152)
  AUTHORS   Tremblay,G.B., Tremblay,A., Copeland,N.G., Gilbert,D.J.,
            Jenkins,N.A., Labrie,F. and Giguere,V.
  TITLE     Cloning, chromosomal localization, and functional analysis of the
            murine estrogen receptor beta
  JOURNAL   Mol. Endocrinol. 11 (3), 353-365 (1997)
  MEDLINE   97211383
```

```
   PUBMED   9058481
REFERENCE   2  (bases 1 to 2152)
  AUTHORS   Tremblay,G.B., Tremblay,A. and Giguere,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1996) Molecular Oncology Group, Royal Victoria
            Hospital, 687 Pine Avenue West, Montreal H3A 1A1, Canada
FEATURES             Location/Qualifiers
     source          1..2152
                     /organism "Mus musculus"
                     /strain "129/Sv"
                     /db_xref "taxon:10090"
                     /chromosome "12"
                     /tissue_type "ovary"
     gene            1..2152
                     /gene "Estrb"
     CDS             12..1469
                     /gene "Estrb"
                     /function "nuclear hormone receptor transcription factor;
                     steroid receptor"
                     /note "ERbeta"
                     /codon_start 1
                     /product "estrogen receptor beta"
                     /protein_id "AAB51142.1"
                     /db_xref "GI:1912468"
                     /translation "MTFYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPTSGHLSPI
                     ATHCQSSLEYAFTQKSPWCFAASLEHTLPVNKETLKEKLGGSQCASPVISPSAKEDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LRKCYEVGMVKGGSRRERCGYRIVRRQRSASEQVHCLNKAKRTSGHIPPVKELLLNSL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEAESSRKLTHLLNAVTDAL
                     VWVISKSGISSQQQSVRLANLTMLTSHVPHTSNKGMEHLISMKCKNVVPVYPLLLEML
                     NAHTLRGYKSSISCSEDCSTEDSKSKECSQNLQSQ"
BASE COUNT      549 a    526 c    570 g    507 t
ORIGIN

  Query Match          58.9%;  Score 1117.4;  DB 10;  Length 2152;
  Best Local Similarity   83.5%;  Pred. No. 1.3e-255;
  Matches 1268;  Conservative    0;  Mismatches  251;  Indels    0;  Gaps    0;

Qy        209 GCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAAC 268
              | ||||||||||| || |||||||||||||| |||||  |||||||||     |  ||||
Db          9 GACATGACATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGGTAAC 68

Qy        269 TTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGG 328
              ||||||||||||| || ||||  |||||||||||| |||||||  ||||
Db         69 CTGGAAGGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTGCTATGGCCAACTTCTGGA 128

Qy        329 CACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCCTCAAAAG 388
              |||||  ||||||||||     |||  |||| | ||| | ||| |||||  ||||||||||||
Db        129 CACCTCTCTCCTTTAGCCACCCACTGCCAATCATCACTTCTCTATGCAGAACCTCAAAAG 188

Qy        389 AGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTG 448
              ||||| ||||||||||||||||| ||||||||||||| ||||||||||||||||||| |||
Db        189 AGTCCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGACCCTG 248

Qy        449 AAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGG 508
              || |||||| || | |||| | |||| |||||||||||||| ||||| |  | ||||||
Db        249 AAGAGGAAGCTTGGCGGGAGCGGTTGTGCCAGCCCTGTTACTAGTCCAAGCGCCAAGAGG 308

Qy        509 GATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGC 568
              |||||||||||||||||| |||||||| ||||| ||||| || ||||| || || ||||||
Db        309 GATGCTCACTTCTGCGCCGTCTGCAGTGATTATGCATCTGGGTATCATTACGGTGTCTGG 368

Qy        569 TCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATT 628
              || ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        369 TCCTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTATATC 428

Qy        629 TGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGC 688
              ||||||||  || |||||||||||  || ||  || ||||||||  || ||||||||||||||
Db        429 TGTCCAGCCACGAATCAGTGTACCATAGACAAGAACCGGCGTAAAAGCTGCCAGGCCTGC 488

Qy        689 CGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGT 748
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 CGACTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGT 548

Qy        749 GGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAG 808
              |||||||| | || || |||||||||||||| ||||||| |||| ||      |||
Db        549 GGGTACCGAATAGTACGAAGACAGAGAAGTGCCAGCGAGCAGGTGCATTGCCTGAACAAA 608

Qy        809 GCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGC 868
              |||||||||    | || ||| | ||||| |||  |||||| |||||| || | |||||
Db        609 GCCAAGAGAACCAGTGGCACACACCCCGGGTGAAGGAGCTACTGCTGAACTCTCTGAGT 668

Qy        869 CCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGC 928
              |||||||||||  |||||||||| ||||| |||||||| ||| |||||||  | |||||
Db        669 CCCGAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTGCTAGTGAGCCGT 728

Qy        929 CCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAG 988
              |||||   |||||||||||||||||||||||||||||||||| || ||| |||| ||||||
Db        729 CCCAGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTCACGAAGCTGGCTGACAAG 788

Qy        989 GAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTG 1048
              ||  |||| ||||||||  |||||||||||||| || || ||||||||||||||||||||
Db        789 GAACTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTG 848

Qy       1049 TTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATG 1108
              || |||||| || |||||||| |||| ||||||||||| | ||| ||||||||||
Db        849 TTGGACCAAGTCCGCCTCTTGGAAAGCTGCTGGATGGAGGTGCTGATGGTGGGGCTGATG 908

Qy       1109 TGGCGGTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGG 1168
              |||||||| || |||||||||||||||||||||||||||||| || ||||||||||||
Db        909 TGGCGGTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGACAGG 968

Qy       1169 CATCAGCGGGAAATCCGTAGAACGAATTCTGCAAATCTTTGACATGCTCCTGGCAACTACT 1228
              |||||||||||| ||||| ||||| ||||||||||||||| ||||||||||| || ||
Db        969 GATGAGGGGAAGTGCGTGGAAGGGATTCTGGAAATCTTTGGCATGCTCCTGGCGACGACG 1028

Qy       1229 TCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATC 1288
              || |||| || |||||||||| || |||||||||||||| ||||| ||||||||||
Db       1029 GCACGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCATGATT 1088

Qy       1289 CTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGATGCTGACAGCAGCCGG 1348
              || ||||| ||||||||||||||  |||  || || | ||||| || || || ||||||
Db       1089 CTCCTCAACTCCAGTATGTACCCCTTGGCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGG 1148

Qy       1349 AAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGC 1408
              |||||| | ||| |  |||||||| ||||| ||||| |||| |||||||||| | |||||
Db       1149 AAGCTGACACACCTATTGAACGCAGTGACAGATGCCCTGGTCTGGGTGATTTCGAAGAGT 1208

Qy       1409 GGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCAC 1468
              || ||||| |||||||||||  ||  | || ||||| ||||||||||||| || || ||
Db       1209 GGAATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTTCTCAT 1268

Qy       1469 GTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTG 1528
              |||||||     |||||||||||||||||||||||||||| ||||||||||||||||||
Db       1269 GTCAGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTG 1328

Qy       1529 GTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAG 1588
              ||||| ||||| |||||||||||||||||||||||||| |||  |||||| |||| ||||
Db       1329 GTCCCGGTGTACGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAG 1388

Qy       1589 TCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCC 1648
              ||||| ||| ||||||| |||||| || || ||||||||||||| ||||||||||||||
Db       1389 TCCTCAATCTCGGGGTCTGAGTGCTGCTCGACAGAGGACAGTAAGAGCAAAGAGGGCTCC 1448

Qy       1649 CAGAACCCACAGTCTCAGTGACGCCTGGCCCTGAGGTGAACTGGCCCACAGAGGTCACAA 1708
              |||||||  ||||| |||||||| |  | |  |||| | |       ||  || |     |
Db       1449 CAGAACCTCCAGTCACAGTGACGGCCAGGCTGGAGGCGGATGATTACAATGATGGTCAGA 1508

Qy       1709 GCTGAAGCGTGAACTCCAG 1727
              ||   | || || | ||
Db       1509 AGTGGGACATGTACCCTAG 1527

RESULT 47
```

```
A61588
LOCUS       A61588                  1458 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 6 from Patent WO9709348.
ACCESSION   A61588
VERSION     A61588.1  GI:3715918
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Kuiper,G.G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348-A 6 13-MAR-1997;
            KAROBIO AB (SE)
FEATURES             Location/Qualifiers
     source          1. .1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT      365 a    373 c    403 g    317 t
ORIGIN

  Query Match             57.9%;  Score 1099.6;  DB 6;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 2.3e-251;
  Matches 1234;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        212 ATGACATTCTATAGGCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTG 271
              ||| ||||||| || |||||||||||||| |||||  |||||||||     | |||| ||
Db          1 ATGGCATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTG 60

Qy        272 GAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCAC 331
              |||||||||||| ||| |||||  ||||||||||||| | |||||||  |||| |||
Db         61 GAAGGTGGGCCTGTTGGCCAGACTGCAAGCCCAAATGTGCTATGGCCAACTTCTGGACAC 120

Qy        332 CTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGT 391
              || ||||||||||    |||  |||| | ||| | ||| ||||| ||||||||||||||
Db        121 CTCTCTCCTTTAGCCACCCAGTGCCAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGT 180

Qy        392 CCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAA 451
              || |||||||||||||| ||||||||||||| ||||||||||||||||| |||||
Db        181 CCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGACCCTCAAC 240

Qy        452 AGGAAGGTTAGTGGCAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGAT 511
              |||||| || | |||| | |||| ||||||||||||||| ||||| |  | |||||||||
Db        241 AGGAAGCTTGGCGGGAGCGGTTGTGCCAGCCCTGTTACTAGTCCAAGCACCAACAGGGAT 300

Qy        512 GCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCG 571
              |||||||||||| || |||||||| ||||| ||||| || ||||| || || |||||||
Db        301 GCTCACTTCTGTGCCGTCTGCAGTGATTATGCATCTGGGTATCATTACGGTGTCTGGTCC 360

Qy        572 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGT 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db        361 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTATATCTGT 420

Qy        632 CCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGA 691
              ||||| || |||||||||| || || || |||||||| || | ||||||||||||||||
Db        421 CCAGCCACGAATCAGTGTACCATAGACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGA 480

Qy        692 CTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGACAGAGATGTCCC 751
              ||||| |||||||||||||| |||||||| |||||||| ||| ||||||| || |||||||
Db        481 CTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGACAAAGGTGTGCC 540

Qy        752 TACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGCCAAGGCC 811
              |||||  | || || |||||||||||||||  ||||||| |||| ||      ||| |||
Db        541 TACCGAATAGTACGAAGACAGAGAAGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCC 600

Qy        812 AAGAGAAGTGGCGGCCACGCGCCCCGGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCC 871
              |||||||   | || ||| | ||||| ||  |||||| |||||| || | ||||| |||
Db        601 AAGAGAACCAGTGGGCACACACCCCGGGTGAAGGAGCTACTGCTGAACCTCTTCGAGTCCC 660

Qy        872 GAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCC 931
              |||||||| |||||||||| ||||| |||||||| ||| |||||||  | || || |||
Db        661 GAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTGCTAGTGAGCGGTCCC 720

Qy        932 AGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAG 991
              ||   |||||||||||||||||||||||||||||||||||||| || ||| |||| |||||||
Db        721 AGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAA 780

Qy        992 TTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTC 1051
              |||| |||||||  |||||||||||||| || || ||||||||||||||||||||||||
Db        781 CTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTG 840

Qy       1052 GACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATCATGGGGCTGATGTGG 1111
              ||||||| || |||||||| ||||| |||||||||||| | ||| ||||||||||||||
Db        841 GACCAAGTCCGGCTCTTGGAAAGCTGCTGGATGGAGGTGTTGATCATGGGGCTGATGTGG 900

Qy       1112 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGAT 1171
              ||||| || |||||||||||||||||||||||||||||||||||| || ||||||||||||
Db        901 CGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGAT 960

Qy       1172 CAGGGCAAATCCGTACAAGCAATTCTCGAAATCTTTGACATGCTCCTGGCAACTACTTCA 1231
              ||||||| |||| |||| |||||||||||||||||||||||||||| || ||  ||
Db        961 GAGGGGAAGTGCGTGGAAGGGATTCTGGAAATCTTTGACATGCTCCTGGCGACGACGGCA 1020

Qy       1232 AGGTTTCGAGATTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1291
              |||| || |||||||||| || |||||||||||||| |||| ||||||||||  ||
Db       1021 CGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCATGATTCTC 1080

Qy       1292 CTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAG 1451
              ||||| ||||||||||||   ||| || || | |||||| || || || |||||||||
Db       1081 CTCAACTCCAGTATGTACCACTTGGCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAG 1140

Qy       1352 CTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGC 1411
              ||| | ||| |  |||||| |||| |||||  |||| |||||||||| | |||||  |
Db       1141 CTGACACACCTATTGAACGCACTGACAGATGCCCTGGTCTGGGTCATTTCGAAGAGTAGA 1200

Qy       1412 ATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTC 1471
              ||||| ||||||||||| ||   | || ||||| ||||||||||||| || || || |||
Db       1201 ATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTC 1260

Qy       1472 AGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTC 1531
              |||||    |||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1261 AGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTC 1320

Qy       1532 CCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAAGTCC 1591
              || ||||| |||||||||||||||||||||||||||| ||| |||||| |||| |||||||
Db       1321 CCGGTGTACGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCC 1380

Qy       1592 TCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAG 1651
              || ||| ||||||| | |||| || || |||||||||||| ||||||||||||||||||
Db       1381 TCAATCTCGGGGTCTGGGTGCTGCTCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAG 1440

Qy       1652 AACCCACAGTCTCAGTGA 1669
              ||||  ||||||||||||
Db       1441 AACCTCCAGTCTCAGTGA 1458


RESULT 38
AR075922
LOCUS       AR075922                1458 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 6 from patent US 5958710.
ACCESSION   AR075922
VERSION     AR075922.1  GI:10002668
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J.-A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710-A 6 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1. .1458
                     /organism="unknown"
BASE COUNT      365 a    373 c    403 g    317 t
ORIGIN
```

```
  Query Match             57.9%;  Score 1099.6;  DB 6;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 2.3e-251;
  Matches 1234;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        212 ATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTG 271
              ||| ||||||| || ||||||||||||||| |||||  |||||||||     | |||| ||
Db          1 ATGGCATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTG 60

Qy        272 GAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCAC 331
              |||||||||||||| ||| ||||| |||||| |||||| | ||||||||  |||| |||
Db         61 GAAGGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTCTATGGCCAACTTCTGGACAC 120

Qy        332 CTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGT 391
              || |||||||||    ||| |||| | ||| | ||| ||||| ||||||||||||||||
Db        121 CTCTCTCCTTTAGCCACCCACTGCCAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGT 180

Qy        392 CCCTGGTGTGCAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAA 451
              || |||||||||||||||| |||||||||||||| ||||||||||||||||| |||||
Db        181 CCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGACCCTGAAG 240

Qy        452 AGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGAT 511
              |||||| || | |||| | |||| |||||||||||||| ||||| |  | ||||||||
Db        241 AGGAAGCTTGGCGGGAGCGGTTGTGCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGAT 300

Qy        512 GCTCACTTCTGCGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCG 571
              |||||||||| || |||||||| ||||| |||| || || ||||| || || ||||||||
Db        301 GCTCACTTCTGTGCCGTCTGCAGTGATTATGCATCTGGGTATCATTACGGTGTCTGGTCC 360

Qy        572 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGT 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |||
Db        361 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTATATCTGT 420

Qy        632 CCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCCTGCCAGCCCTGCCGA 691
              ||||| || |||||||||||| || || || ||||||| || | ||||||||||||||||
Db        421 CCAGCCACGAATCAGTGTACGATAGACAAGAACCGGCGTAAAAACCTGCCAGGCCTGCCGA 480

Qy        692 CTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGG 751
              ||||| |||||||||||||| |||||||| |||||||| ||| ||||||| || ||||||
Db        481 CTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGG 540

Qy        752 TACCGCCTTGTGCGGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCCGCAAGCCC 811
              |||||  | || || |||||||||||||||  | ||||| |||| ||      ||| ||||
Db        541 TACCGAATAGTACGAAGACAGAGAAGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCC 600

Qy        812 AAGAGAAGTGGCCGCCACCGCCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCC 871
              |||||||  | || || | |||| ||| |||||| |||||| || | |||| |||
Db        601 AAGAGAACCAGTGGGCACACACCCCGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCC 660

Qy        872 GAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGACACGCCGCCC 931
              |||||||| |||||||||| |||| |||||||| ||| |||||||  | || || |||
Db        661 GAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTGCTAGTCAGTCGTCCC 720

Qy        932 AGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAG 991
              ||  |||||||||||||||||||||||||||||||| || ||| |||| ||||||||
Db        721 AGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAA 780

Qy        992 TTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTC 1051
              |||| ||||||||  |||||||||||| || || ||||||||||||||||||||||||||
Db        781 CTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTG 840

Qy       1052 GACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGG 1111
              |||||||| || |||||||| |||| ||||||||||| | ||| ||||||||||||||||
Db        841 GACCAAGTCCGCCTCTTGGAAAGCTGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGG 900

Qy       1112 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGAT 1171
              ||||| || |||||||||||||||||||||||||||||||| || ||||||||||||||||
Db        901 CGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGAT 960

Qy       1172 GAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCA 1231
              |||||||| |||| |||| |||||||||||||||||||||||||||| || ||  ||
Db        961 GAGGGAAGTGCGTGGAAGGGATTCTGGAAATCTTTGACATGCTCCTGGCGACGACGGCA 1020

Qy        1232 AGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1291
               |||| || |||||||||| || |||||||||||||| ||||| ||||||||||| ||
Db        1021 CGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCATGATTCTC 1080

Qy        1292 CTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAG 1351
               ||||| |||||||||||||   ||| || || | |||||| || || || |||||||||
Db        1081 CTCAACTCCAGTATGTACCACTTGGCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAG 1140

Qy        1352 CTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGC 1411
               ||| | ||| |  |||||| ||||| |||||  |||| |||||||||| | |||||  |
Db        1141 CTGACACACCTATTGAACGCAGTGACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGA 1200

Qy        1412 ATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTC 1471
               ||||| |||||||||||  ||  | || |||| ||||||||||| || || || |||
Db        1201 ATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTC 1260

Qy        1472 AGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTC 1531
               |||||    |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        1261 AGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTC 1320

Qy        1532 CCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCC 1591
               || ||||| |||||||||||||||||||||||||||||| |||  |||||| |||| ||||
Db        1321 CCGGTGTACGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCC 1380

Qy        1592 TCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAG 1651
               || ||| ||||||| | |||| || || ||||||||||||| ||||||||||||||||||
Db        1381 TCAATCTCGGGGTCTGGGTGCTGCTCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAG 1440

Qy        1652 AACCCACAGTCTCAGTGA 1669
               ||||  ||||||||||||
Db        1441 AACCTCCAGTCTCAGTGA 1458


RESULT 39
AX022460
LOCUS       AX022460               1458 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 6 from Patent EP0935000.
ACCESSION   AX022460
VERSION     AX022460.1  GI:10046070
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT      365 a    373 c    403 g    317 t
ORIGIN

  Query Match             57.9%;  Score 1099.6;  DB 6;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 2.3e-251;
  Matches 1234;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

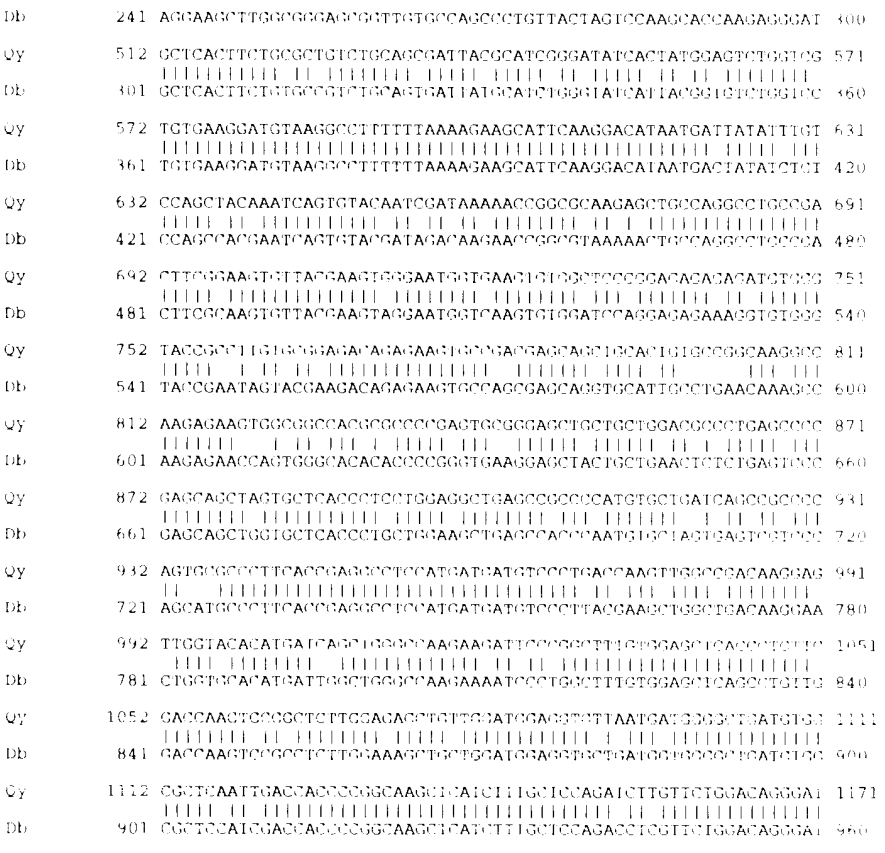
Qy        212 ATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTG 271
              ||| ||||||| || ||||||||||||||| |||||  |||||||||     | |||| ||
Db          1 ATGGCATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTG 60

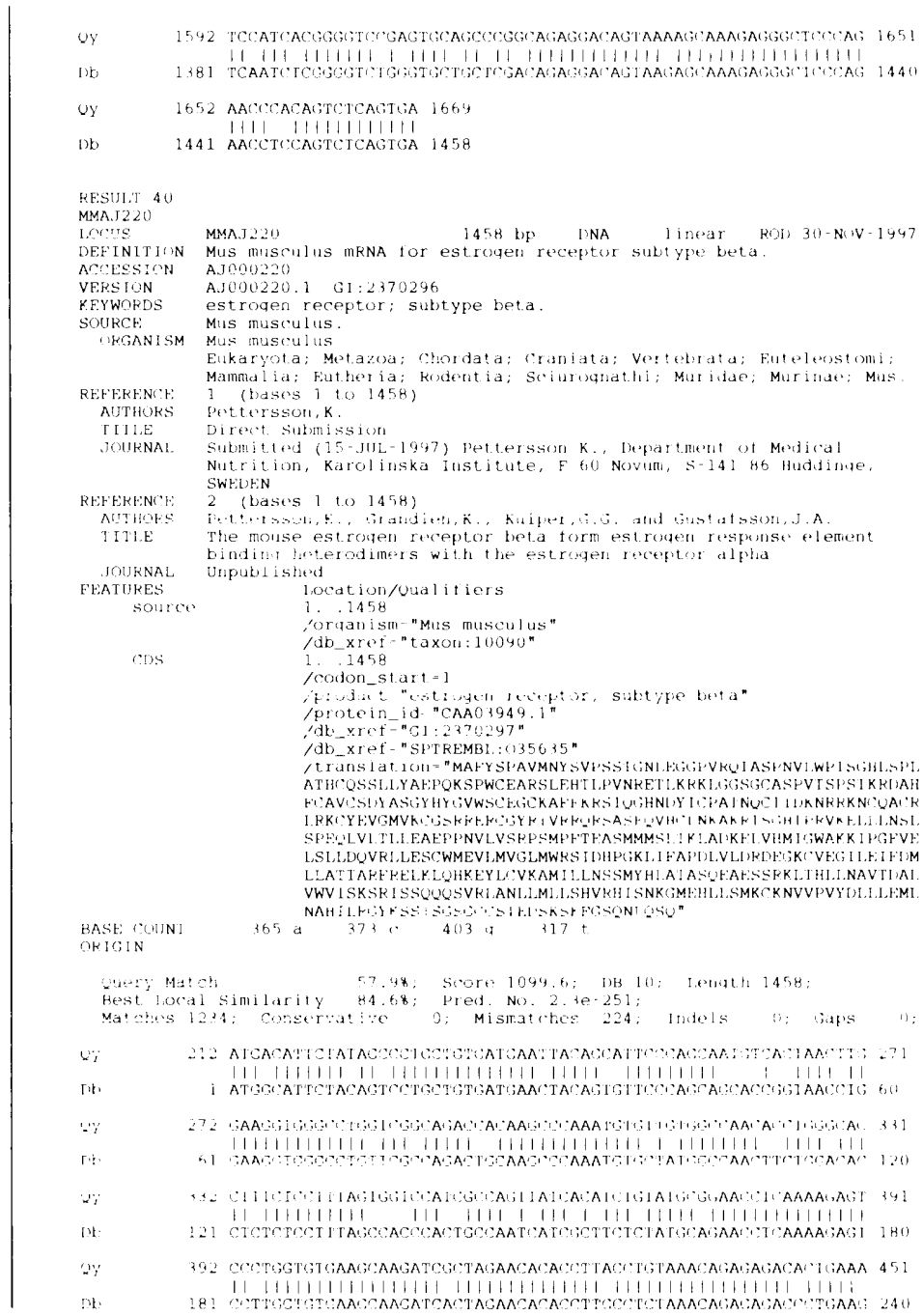
Qy        272 GAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCAC 331
              |||||||||||||| ||| ||||| |||||| |||||| | ||||||||  |||| |||
Db         61 GAAGGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTCTATGGCCAACTTCTGGACAC 120

Qy        332 CTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGT 391
              || |||||||||    ||| |||| | ||| | ||| ||||| ||||||||||||||||
Db        121 CTCTCTCCTTTAGCCACCCACTGCCAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGT 180

Qy        392 CCCTGGTGTGCAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAA 451
              || |||||||||||||||| |||||||||||||| ||||||||||||||||| |||||
Db        181 CCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGACCCTGAAG 240

Qy        452 AGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGAT 511
              |||||| || | |||| | |||| |||||||||||||| ||||| |  | ||||||||
```





```
RESULT 40
MMAJ220
LOCUS       MMAJ220                 1458 bp    DNA     linear   ROD 30-NOV-1997
DEFINITION  Mus musculus mRNA for estrogen receptor subtype beta.
ACCESSION   AJ000220
VERSION     AJ000220.1  GI:2370296
KEYWORDS    estrogen receptor; subtype beta.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Pettersson,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1997) Pettersson K., Department of Medical
            Nutrition, Karolinska Institute, F 60 Novum, S-141 86 Huddinge,
            SWEDEN
REFERENCE   2  (bases 1 to 1458)
  AUTHORS   Pettersson,K., Grandien,K., Kuiper,G.G. and Gustafsson,J.A.
  TITLE     The mouse estrogen receptor beta form estrogen response element
            binding heterodimers with the estrogen receptor alpha
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             1..1458
                     /codon_start=1
                     /product="estrogen receptor, subtype beta"
                     /protein_id="CAA03949.1"
                     /db_xref="GI:2370297"
                     /db_xref="SPTREMBL:O35645"
                     /translation="MAFYSPAVMNYSVPSSIGNLEGGPVRQIASPNVILWPISGHLSPL
                     ATHCQSSLLYAEPQKSPWCEARSLEHTLPVNRETLKRKLGGSGCASPVISPSIKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACR
                     LRKCYEVGMVKGGSKRFEPCGYRKVKRQRSASPQVHCLNKAKRISGHLPGRVKELLLNSL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTPASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATSARFRELKLQHKEYLCVKAMILLNSSMYPLVTATQSEAESSRKLTHLLNAVTDAL
                     VWVISKSRISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHILEGYSSEISGSQCGSTEDSSKSPEGSQNLQSQ"
BASE COUNT      365 a    373 c    403 g    317 t
ORIGIN

  Query Match          57.9%;  Score 1099.6;  DB 10;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 2.3e-251;
  Matches 1294;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        212 ATGACATTCTATAGCCCCTGCTGTGATGAACTACAGCATTCCCACCAAATGTCACTAACTTCT 271
              ||| ||||||| || |||||||||||||| ||||| |||||||||    | |||| ||
Db          1 ATGGCATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCCGTAACCTG 60

Qy        272 GAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCTAACACCTGGGCAC 331
              |||||||||||| ||| |||||  |||||||||||| | |||||||  |||| |||
Db         61 GAAGCTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTGCTATGGCTAACTTCTGCACAC 120

Qy        332 CTTTCTCCTTTAGTGGTCCATGCCAGTTATCACATCTGTATGCGAACCTCAAAAGAGT 391
              || ||||||||||    ||| |||| | ||| | ||| ||||| ||||||||||||||
Db        121 CTCTCTCCTTTAGCCACCCACTGCCAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGT 180

Qy        392 CCCTGGTGTGAAGCAAGATCGCTAGAACACAGCTTAGCTGTAAACAGAGAGACACTGAAA 451
              || |||||||||||||||| |||||||||||||| ||||||||||||||||| |||||
Db        181 CCTTGCTGTGCAACCAACATCACTAGAACACACGTTGCCTCTAAACAGACACACCCTGAAG 240
```

```
Qy        452 AGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGAT 511
              ||||||  || | |||| | |||| ||||||||||||||| |||||  |  | |||||||||
Db        241 AGGAAGCTTGGCGGGAGCGGTTGTGCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGAT 300

Qy        512 GCTCACTTCTGCGGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCG 571
              |||||||||||| || |||||||| |||||  |||||  || |||||  || ||  ||||||||
Db        301 GCTCACTTCTGTGCCGTCTGCAGTGATTATGCATCTGGGTATCATTACGGTGTCTGGTCC 360

Qy        572 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGT 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||  |||
Db        361 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTATATCTGT 420

Qy        632 CCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGA 691
              |||||  || ||||||||||| || ||  || |||||||  || |  ||||||||||||||||
Db        421 CCAGCCACGAATCAGTGTACGATACGACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGA 480

Qy        692 CTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGG 751
              |||||  |||||||||||||||  ||||||||  ||||||||  |||  |||||||  ||  ||||||
Db        481 CTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGG 540

Qy        752 TACCGCCTTGTGCCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCC 811
              |||||  | ||  ||  ||||||||||||||  |||||||  ||||  ||       |||  |||
Db        541 TACCGAATAGTACGAAGACAGAGAAGTGCCAGCCAGCAGGTGCATTGCCTGAACAAAGCC 600

Qy        812 AAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCC 871
              |||||||   | ||  |||  | |||||  |||  ||||||  ||||||  ||  |  |||||  |||
Db        601 AAGAGAACCAGTGGGCACACACCCCGGGTGAAGGAGCTACTGCTGAACTCTCTTGAGTCCC 660

Qy        872 GAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCC 931
              |||||||| ||||||||||| |||||  |||||||| |||  |||||||   | ||  ||  |||
Db        661 GAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCC 720

Qy        932 AGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAG 991
              ||   |||||||||||||||||||||||||||||||||||| ||  |||  ||||  ||||||||
Db        721 AGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAA 780

Qy        992 TTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTC 1051
              ||||  ||||||||  ||||||||||||| || ||  |||||||||||||||||||||||
Db        781 CTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTG 840

Qy       1052 GACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGG 1111
              ||||||||  || ||||||||  ||||  ||||||||||||  |  |||  ||||||||||||||
Db        841 GACCAAGTCCGCCTCTTGGAAAGCTGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGG 900

Qy       1112 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGAT 1171
              |||||  || |||||||||||||||||||||||||||||||||||||  ||  ||||||||||||||
Db        901 CGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGAT 960

Qy       1172 GAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCA 1231
              ||||||||  |||||  |||||  |||||||||||||||||||||||||||||||  ||  ||  ||
Db        961 GAGGGGAAGTGCGTGGAAGGCATTCTGGAAATCTTTGACATGCTCCTGGCGACGACGGCA 1020

Qy       1232 AGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1291
              ||||  ||  ||||||||||  ||  ||||||||||||||  |||||  ||||||||||  ||
Db       1021 CGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCATGATTCTC 1080

Qy       1292 CTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAG 1351
              |||||  ||||||||||||||    |||  ||  ||  |  ||||||  ||  ||  ||  ||||||||
Db       1081 CTCAACTCCAGTATGTACCACTTGGCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAG 1140

Qy       1352 CTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGC 1411
              |||  |  |||  |  |||||||  |||||  ||||  ||||  |||||||||  |  |||||  |
Db       1141 CTGACACACCTATTGAACGCAGTGACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGA 1200

Qy       1412 ATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTGCTGTCCCACGTC 1471
              |||||  ||||||||||||  ||  |  ||  |||||  |||||||||||||  ||  ||  |||
Db       1201 ATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTC 1260

Qy       1472 AGGCATGCGAGTAACAAAGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTC 1531
              |||||    |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1261 AGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTC 1320

Qy       1532 CCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAAGTCC 1591
              ||  |||||  |||||||||||||||||||||||||||||||  |||  ||||||  ||||  ||||||
Db       1321 CCGGTGTACGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCC 1380

Qy       1592 TCCATCACGGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAG 1651
              ||  |||  |||||||  |  ||||  ||  ||  ||||||||||||||  ||||||||||||||||||
Db       1381 TCAATCTCGGGGTCTGGGTGCTGCTCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAG 1440

Qy       1652 AACCCACAGTCTCAGTGA 1669
              ||||  ||||||||||||
Db       1441 AACCTCCAGTCTCAGTGA 1458
```

Search completed: June 16, 2003, 07:49:03
Job time : 5033.77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 23:19:18 ; Search time 435.175 Seconds
(without alignments)
9822.009 Million cell updates/sec

Title: US-08-826-361A-24
Perfect score: 1898
Sequence: 1 CACGAATCTTTGAGAACATT.....TGAATCCAGCCTTTCCACAC 1898

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcqdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1898	100.0	1898	24	AAI72144	Full length estrog
2	1898	100.0	1936	18	AAT88415	Human oestrogen re
3	1718	90.5	1749	22	AAF59897	Human oestrogen re
4	1683	88.7	2433	24	ABN86125	Nucleotide sequenc
5	1669	87.9	2011	22	AAD18442	Human estrogen rec
6	1669	87.9	2011	22	AAS43105	Human cDNA encodin
7	1667.4	87.9	1686	20	AAX24364	Human oestrogen re
8	1635.6	86.2	1647	20	AAX34309	Human oestrogen re
9	1453.6	76.6	1460	18	AAT62843	Human oestrogen re
10	1450.4	76.4	1460	20	AAX26686	DNA encoding an es
11	1434	75.6	1434	18	AAT88412	Human oestrogen re
12	1134	75.6	1434	24	AAI72128	Estradiol receptor
13	1256.4	66.2	1820	21	AAZ89597	Bovine ER beta cDN
14	1247	65.7	1251	18	AAT88413	Human oestrogen re
15	1247	65.7	1251	24	AAI72153	cDNA encoding ER s
16	1247	65.7	1257	18	AAT88414	Human oestrogen re
17	1247	65.7	1257	24	AAI72146	cDNA encoding ER s
18	1239.6	65.3	2568	18	AAT62842	Rat oestrogen rece
19	1239.6	65.3	2568	20	AAX26685	DNA encoding an es
20	1160.2	61.1	1704	20	AAX33382	Murine mER beta-3
21	1158.6	61.0	1704	20	AAX24935	Mouse complete oes
22	1099.6	57.9	1458	18	AAT62844	Mouse oestrogen re
23	583	30.7	640	22	AAF59900	EGFP/human oestrog
24	438	23.1	597	22	ABA61882	Human foetal liver
25	438	23.1	597	22	ABA29447	Probe #7903 for qe
26	438	23.1	597	22	AAK10196	Human brain expres
27	438	23.1	597	22	AAK36093	Human bone marrow
28	438	23.1	597	22	AAI41808	Probe #18494 used
29	438	23.1	597	24	ABS10245	Human genome-deriv
30	438	23.1	425791	22	AAS43104	Human oestrogen re
31	395.8	20.9	2764	24	AAD21046	Medaka fish oestro
32	390.8	20.6	1374	22	AAF29925	Human estrogen rec
33	390.8	20.6	1788	24	ABL57407	Human nuclear oest
34	390.8	20.6	4963	16	AAT06873	PhCMV*-1 promoter
35	390.8	20.6	4963	16	AAQ76270	PhCMV* 1 controlle
36	390.8	20.6	4963	20	AAX81725	cDNA encoding rabb
37	390.8	20.6	4963	20	AAX27906	Rabbit progesteron
38	390.8	20.6	4963	20	AAX01370	PhCMV*-1 rabbit pr
39	390.8	20.6	4963	22	AAD09838	Rabbit progesteron
40	389.8	20.5	1824	22	AAI68126	Lepomis contrarchi
41	389.8	20.5	3460	22	AAI68124	Lepomis contrarchi
42	389.4	20.5	1728	21	AAA92174	Oryzias lapites oe
43	389.4	20.5	1863	21	AAA92175	Oryzias lapites oe
44	389.2	20.5	2092	22	AAC86920	Nucleotide sequenc
45	389.2	20.5	2092	24	ABL51900	Human oestrogen re
46	389.2	20.5	6460	20	AAZ23433	Human oestrogen rec
47	387.6	20.4	2092	8	AAN70880	cDNA encoding huma
48	381	20.1	1965	24	ABA98661	Oestrogen receptor
49	375.2	19.8	1223	22	AAF29924	Human estrogen rec
50	373.8	19.7	1765	24	ABQ72931	Caiman crocodilus
51	351.8	18.5	1746	24	ABN81366	Cnemidophorus unip
52	337.4	17.8	6954	22	AAD06053	Plasmid C7LBDCS en
53	337.4	17.8	6956	22	AAD06052	Plasmid C7LBDCL en
54	337.4	17.8	7038	22	AAD06046	Plasmid 2C7LBDCS e
55	319.2	16.8	996	22	AAI68125	Lepomis oestrogen
56	306	16.1	306	22	ABA74377	Human foetal liver
57	306	16.1	306	20	ABA39287	Probe #17753 for q
58	306	16.1	306	22	AAK22844	Human brain expres
59	306	16.1	306	22	AAK49015	Human bone marrow
60	306	16.1	306	22	AAI54843	Probe #24529 used
61	306	16.1	306	24	ABS22646	Human genome-deriv
62	302	15.9	586	22	AAT42100	Probe #10786 used
63	300	15.8	300	22	AAI55133	Probe #24819 used
64	288.8	15.2	11365	24	ABK87505	Directed evolution
65	288.2	15.2	1380	22	AAC86919	Nucleotide sequenc
66	287.2	15.1	1446	22	AAC88199	Mammalian two-hybr
67	287.2	15.1	6567	22	AAD06054	Plasmid E2C7LBDAS e
68	287.2	15.1	6624	22	AAD06049	Plasmid C7LBDAS en
69	287.2	15.1	6639	22	AAD06055	Plasmid E2C7LBDBS e
70	287.2	15.1	6695	22	AAD06051	Plasmid C7LBDBS en
71	287.2	15.1	6746	22	AAD06048	Plasmid C7LBDAL en
72	287.2	15.1	6801	22	AAD06056	Plasmid LBDASNLSVP
73	287.2	15.1	6801	22	AAD06059	Plasmid C7LBDBSVPl
74	287.2	15.1	6819	22	AAD06050	Plasmid C7LBDBL en
75	287.2	15.1	6828	22	AAD06044	Plasmid 2C7LBDAS e
76	287.2	15.1	6930	22	AAD06045	Plasmid 2C7LBDBS e
77	285.6	15.0	1956	19	AAV17756	Multi chimeric tra
78	285.6	15.0	1956	19	AAV18690	Multi-chimeric tra
79	285.6	15.0	2220	24	ABA31104	Fused ER protein -
80	285.6	15.0	2322	20	AAX60628	SSR-LBD fusion pol
81	285.6	15.0	6695	22	AAD06057	Plasmid C7LBDBSG40
82	285.6	15.0	6695	22	AAD06058	Plasmid C7LBDBSG52

83	284	15.0	1984	24	ABL57498	Cre recombinase-mu
84	280.8	14.8	1983	24	ABL57500	Cre recombinase-mu
85	279.2	14.7	1983	24	ABL57499	Cre recombinase-mu
86	271.8	14.3	1047	24	ABK87184	Endocrine disrupti
87	265.4	14.0	1551	22	AAD06060	VP16?ER construct
88	265.4	14.0	2749	24	AAI72479	E2C EE L ER VP64 c
89	265.4	14.0	2775	24	AAI72480	E2C ER LL ER VP64
90	232.4	12.2	599	22	ABA59744	Human foetal liver
91	232.4	12.2	599	22	ABA28252	Probe #6218 for qe
92	232.4	12.2	599	22	AAK08012	Human brain expres
93	232.4	12.2	599	22	AAK43884	Human bone marrow
94	232.4	12.2	599	22	AAI39610	Probe #8296 used t
95	232.4	12.2	599	24	ABS08715	Human genome deriv
96	183	9.6	183	22	ABA72285	Human foetal liver
97	183	9.6	183	22	ABA38142	Probe #16598 for q
98	183	9.6	183	22	AAK20707	Human brain expres
99	183	9.6	183	22	AAK46853	Human bone marrow
100	183	9.6	183	22	AAI52692	Probe #21378 used
101	183	9.6	183	24	ABS21150	Human genome-deriv
102	157.6	8.3	549	20	AAX37813	pDM16 oestrogen re
103	157	8.3	521	20	AAX37808	pDM12 oestrogen re
104	157	8.3	566	20	AAX37812	pDM14 oestrogen re
105	154.8	8.1	1470	22	AAS06086	Angiotensin conver
106	154.8	8.1	1640	22	AAS06058	Angiotensin conver
107	142.2	7.5	204	16	AAQ85976	Human estrogen rec
108	133.8	7.0	2218	24	ABK83889	Human cDNA differe
109	133.8	7.0	2218	24	ABL64390	Stomach cancer rel
110	129	6.8	2249	9	AAN80921	Sequence encoding
111	129	6.8	2539	22	AAI58567	Human polynucleoti
112	129	6.8	4250	24	ABA92902	Human nuclear horm
c 113	121.2	6.4	2559	22	AAI60353	Human polynucleoti
114	113.8	6.0	1377	21	AAA40078	Human ERRgamma cod
115	113.8	6.0	1441	21	AAZ95741	Human oestrogen re
116	113.8	6.0	1615	19	AAV47645	Steroid hormone re
117	113.8	6.0	2985	20	AAX32266	Human nuclear rece
118	113.8	6.0	2985	21	AAA09801	Human nuclear rece
119	113.8	6.0	2987	20	AAX32267	Human nuclear rece
120	113.8	6.0	2987	21	AAA09802	Human nuclear rece

ALIGNMENTS

```
RESULT 1
AAI72144
ID   AAI72144 standard; cDNA; 1898 BP.
XX
AC   AAI72144;
XX
DT   25-MAR-2002  (first entry)
XX
DE   Full length estrogen receptor cDNA.
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER-alpha; ER-beta; exon 8;
KW   estradiol; nuclear receptor; progesterone receptor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..1669
FT                   /*tag= a
FT                   /product= "Full length ER"
XX
PN   EP1162264-A2.
XX
PD   12-DEC-2001.
XX
PF   25-MAR-1997; 2001EP-0202021.
XX
PR   26-MAR-1996;  96EP-0200820.
PR   22-NOV-1996;  96EP-0203284.
PR   25-MAR-1997;  97EP-0200903.
XX
PA   (ALKO ) AKZO NOBEL NV.
XX
PI   Mosselman S, Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   P-PSDB; AAB47835.
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor
XX
PS   Example A; Page 26; 35pp; English.
XX
CC   This sequence encodes a full length novel estrogen receptor (ER). The
CC   additional 5' and 3' sequences of this sequence, compared to the ER
CC   sequence given in AAI72128, were generated using RACE PCR. The gene
CC   encoding this new ER is located on chromosome 14 and has a different
CC   tissue distribution from classical ER. This ER also has two orphan ER's,
CC   ER-alpha and ER-beta. These orphan receptors have estrogen receptor
CC   related structure but do not appear to able to bind estradiol or other
CC   ER ligands. The DNA binding domain (DBD) and ligand binding domain (LBD)
CC   from this ER may be used in the chimeric receptor of the invention
CC   which also has an N-terminal domain. The chimeric receptor, or DNA
CC   encoding it, is useful in a screening assay for identification of new
CC   drugs. Similar chimeric receptors comprising the LBD of the new ER,
CC   and also comprising the DBD and an N-terminal domain derived from
CC   another nuclear receptor e.g., progesterone receptor, are useful for the
CC   screening of compounds to identify new ligands or hormone analogs which
CC   are able to activate the new ER. Chimeric receptors comprising a DBD
CC   of the new ER, and LBD and an N-terminal domain derived from
CC   another nuclear receptor, can be used to identify new ligands or
CC   hormone analogs for the nuclear receptors.
XX
SQ   Sequence 1898 BP; 452 A; 506 C; 501 G; 439 T; 0 other;

  Query Match          100.0%;  Score 1898;  DB 24;  Length 1898;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1898;  Conservative    0,  Mismatches    0,  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATGGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATGGCTAGAACA 420

Qy        421 CACTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
```

```
Db        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720

Qy        721 GAAGTGTGGCTCCCGGAGAGAGACATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAGTGTGGCTCCCGGAGAGAGACATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCGAGT 840

Qy        841 GCGGGAGGTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGGGAGGTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCCAAGAA 1020

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTC 1080

Qy       1081 GATGCAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATGCAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGCA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGCA 1200

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTTCGAGAGTTAAAACTCCAACACAA 1260

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCCGTGACCGA 1380

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAACAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560

Qy       1561 GAATGCCCACGTGCTTCGCGGGTCCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAATGCCCACGTGCTTCGCGGGTCCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680

Qy       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCCTGAACTCCAGTGTGTCAGGAGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCCTGAACTCCAGTGTGTCAGGAGCC 1740

Qy       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800

Qy       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTGCCCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTGCCCT 1860

Qy       1861 TGTTAGCAGAGGGACATTTGAATCGAGCGTTTCCACAC 1898
              |||||||||||||||||||||||||||||||||||||||
Db       1861 TGTTAGCAGAGGGACATTTGAATCGAGCGTTTCCACAC 1898
```

```
RESULT 2
AAT88415
ID   AAT88415 standard; cDNA; 1936 BP.
XX
AC   AAT88415;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human oestrogen receptor protein cDNA #2.
XX
KW   Oestrogen receptor protein; steroid; alternative splicing; estradiol;
KW   estone; estriol; screening; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..1669
FT                   /*tag= a
FT                   /product= oestrogen receptor
XX
PN   EP798378-A2.
XX
PD   01-OCT-1997.
XX
PF   25-MAR-1997;   97EP-0200903.
XX
PR   22-NOV-1996;   96EP-0203284.
PR   26-MAR-1996;   96EP-0200820.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Dijkema R,  Mosselman S;
XX
DR   WPI; 1997-473188/44.
DR   P-PSDB; AAW33215.
XX
PT   DNA encoding estrogen receptor  useful in screening assay to
PT   identity novel ligands or hormonal analoques
XX
PS   Claim 5; Page 33-35; 45pp; English.
XX
CC   This sequence encodes a novel oestrogen binding protein isolated from
CC   human testis cDNA in order to study upstream translation-initiation
CC   codons using 5' RACE-PCR technology. This receptor is able to bind and
CC   be activated by estradiol, estone and estriol, can be used in a screening
CC   assay for the identification of new drugs e.g. novel ligands or hormonal
CC   analoques.
XX
SQ   Sequence 1936 BP; 462 A; 511 C; 511 G; 449 T; 0 other;
```

```
  Query Match          100.0%;  Score 1898;  DB 18;  Length 1936;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1898;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCGCGCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGAGCCGCCCCATGTGCTGATCAGCGCGCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960

Qy        961 CATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680

Qy       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGGAGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGGAGCC 1740

Qy       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TGGGCTTCATCTTTCTGCTGTGTGGTCCCTCATTTGGTGATGGCAGGCTTGGTCATGTAC 1800

Qy       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTTCCCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CATCCTTCCCTCCACCTTCCCAACTCTCAGGAGTCGGTGTGAGGAAGCCATAGTTTCCCT 1860

Qy       1861 TGTTAGCAGAGGGACATTTGAATCGAGCGTTTCCACAC 1898
              ||||||||||||||||||||||||||||||||||||||
Db       1861 TGTTAGCAGAGGGACATTTGAATCGAGCGTTTCCACAC 1898


RESULT 3
AAF59897
ID   AAF59897 standard; cDNA; 1740 BP.
XX
AC   AAF59897;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Human oestrogen receptor beta (ER beta) cDNA, SEQ ID NO:1.
XX
KW   Human oestrogen receptor beta; ER beta; labelled hormone receptor;
KW   optical signal; ligand binding; hormonal disorder;
KW   sex hormone associated, ovulation, drug screening, drug discovery, ss.
```

```
XX
OS   Homo sapiens.
XX
PN   WO200107919-A1.
XX
PD   01-FEB-2001.
XX
PF   24-JUL-2000; 2000WO-JP04930.
XX
PR   23-JUL-1999;   99JP-0209860
PR   31-MAY-2000; 2000JP-0163475.
PR   31-MAY-2000; 2000JP-0163476.
XX
PA   (OLYU ) OLYMPUS OPTICAL CO LTD.
XX
PI   Sakamoto H,  Kato N;
XX
DR   WPI; 2001-168581/17.
DR   P-PSDB; AAB60649.
XX
PT   Determination of the interaction of a substance for investigation and
PT   diagnosis of hormonal disorders using an optically labelled receptor
PT   protein -
XX
PS   Claim 26; Page 68-73; 75pp; Japanese.
XX
CC   The invention relates to a method for examining the interaction of a
CC   test substance with a hormone receptor protein. The method involves
CC   contacting a test substance with a hormone receptor protein (e.g., an
CC   oestrogen receptor) which is labelled with an optical marker (e.g., a
CC   fluorescent protein), where the marker is capable of undergoing a change
CC   in its optical properties when the receptor binds a ligand. The optical
CC   signal generated in the presence of the test compound is then compared
CC   with that generated in the absence of the test compound to determine
CC   whether the test compound is interacting with the hormone receptor
CC   protein. The invention also encompasses the labelled hormone receptor
CC   protein, the gene encoding it, and vectors and host cells comprising
CC   such genes. The method of the invention is used for the investigation
CC   and diagnosis of hormonal disorders, especially those associated with
CC   sex hormones e.g., suppression of ovulation. The present sequence
CC   represents cDNA encoding human oestrogen receptor beta (ER-beta).
XX
SQ   Sequence 1740 BP; 425 A; 463 C; 459 G; 393 T; 0 other;

  Query Match             90.5%;  Score 1718;  DB 22;  Length 1740;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1718;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 82

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 142

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 202

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATCAA 262

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 322

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 382

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 442

Qy        421 CACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 CACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 502

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 562

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 622

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 682

Qy        661 AAACCGGCCCAAGAGTTGCCAGGCTGCCGACTTCGGAAGTGTTACGAAGTGGCAATCGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 AAACCGGCCAAGAGTTGCCAGGCTGCCGACTTCGGAAGTGTTACGAAGTGGCAATCGT 742

Qy        721 CAAGTGTGGCTCCCGCAGACAGAGATGTGGGTACCGCCTTGTGCCAGACAGAGAACTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCCAGACAGAGAAGTGC 802

Qy        781 CGACGAGCAGCTGCACTGTGCCGCCAAGGCCAAGAGAAGTGGCGGGCCACGGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CGACGAGCAGCTGCACTGTGCCGCCAAGGCCAAGAGAAGTGGCGGGCCACGGCCCCGAGT 862

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 922

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCCCCTTCACCGAGGCCTCCATGAT 982

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1042

Qy       1021 GATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 GATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1102

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1162

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGA 1222

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1223 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1282

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1283 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1342

Qy       1321 AGCCACCCAGGATGCTCACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1343 AGCGACCCAGGATGCTCACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1402

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1403 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1462

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1463 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1522

Qy        1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1523 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1582

Qy        1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1583 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1642

Qy        1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1643 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1702

Qy        1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1718
               ||||||||||||||||||||||||||||||||||||||
Db        1703 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1740


RESULT 4
AB086125
ID   AB086125 standard; DNA; 2433 BP.
XX
AC   AB086125;
XX
DT   23-SEP 2002  (first entry)
XX
DE   Nucleotide sequence of a fusion GFP/hER beta gene.
XX
KW   Fluorometric analyzer; laser; fluorescence; GFP; ER beta; human; gene;
KW   green fluorescent protein; estrogen receptor beta; fusion; ds.
XX
OS   Synthetic.
OS   Aequorea victoria.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             8..2389
FT                   /*tag= a
FT                   /product= "GFP/hER beta fusion protein"
XX
PN   WO200248693-A1.
XX
PD   20-JUN-2002.
XX
PF   14-DEC-2001; 2001WO-JP10998.
XX
PR   14-DEC-2000; 2000JP-0380827.
PR   30-JAN-2001; 2001JP-0022105.
XX
PA   (OLYU ) OLYMPUS OPTICAL CO LTD.
XX
PI   Kato N, Sakamoto H;
XX
DR   WPI; 2002-528176/56.
DR   P-PSDB; AB880756.
XX
PT   Autocorrection function embedded confocal optics-based fluorometric
PT   analyzers for studying behaviors of fluorescence-labeled molecules
PT   particularly intracellular biological molecules like proteins at
PT   molecular level
XX
PS   Example 2; Page 98-105; 109pp; Japanese.
XX
CC   The invention relates to a fluorometric analyzer that comprises a laser
CC   light source, an optical system for focusing the light beam from laser
CC   light source to a sample to form a confocal region, another optical
CC   system for focusing fluorescence from the sample, a light detector for
CC   measuring intensity, and a recording means to record the variation of
CC   intensity in the measured fluorescence. The analyzers are for studying
CC   behaviors of fluorescence-labeled molecules particularly intracellular
CC   biological molecules like proteins at molecular level, e.g. protein
CC   functions and interactions. Such analyzers are stable and convenient to
CC   operate, thereby enabling easy performance of fluorescence correlation
CC   spectroscopy, fluorescence intensity distribution analysis and
CC   fluorescence intensity multiple distribution analysis. The present
CC   sequence represents the nucleotide sequence of a fusion gene comprising
CC   a GFP (green fluorescent protein) gene and hER (human estrogen receptor)
CC   beta gene.
XX
SQ   Sequence 2433 BP; 587 A; 704 C; 659 G; 483 T; 0 other;

  Query Match             88.7%;  Score 1683;  DB 24;  Length 2433;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1683;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          31 TTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAAAAA 90
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         751 TTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAAAAA 810

Qy          91 CTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACCCCT 150
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         811 CTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACCCCT 870

Qy         151 GGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCCAGC 210
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         871 GGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCCAGC 930

Qy         211 CATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTT 270
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         931 CATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTT 990

Qy         271 GGAAGGTGGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCA 330
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         991 GGAAGGTGGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCA 1050

Qy         331 CCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAG 390
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1051 CCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAG 1110

Qy         391 TCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAA 450
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1111 TCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAA 1170

Qy         451 AAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 510
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1171 AAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 1230

Qy         511 TGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTC 570
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1231 TGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTC 1290

Qy         571 GTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTG 630
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1291 GTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTG 1350

Qy         631 TCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCG 690
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1351 TCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCG 1410

Qy         691 ACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCCTCCCGGAGAGAGAGATGTGG 750
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1411 ACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGG 1470

Qy         751 GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGC 810
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1471 GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGC 1530

Qy         811 CAACAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCC 870
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1531 CAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCC 1590

Qy         871 CGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCC 930
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1591 CGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCC 1650
```

```
Qy        931 CAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGCA 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1651 CAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGCA 1710

Qy        991 GTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTT 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1711 GTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTT 1770

Qy       1051 CGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGCTGATCTC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1771 CGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGCTGATCTC 1830

Qy       1111 GCGCTCAATTGACCACCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGA 1170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1831 GCGCTCAATTGACCACCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGA 1890

Qy       1171 TGAGGGGAAATGCGTAGAAGCAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTC 1230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1891 TGAGGGGAAATGCGTAGAAGCAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTC 1950

Qy       1231 AAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCT 1290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1951 AAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCT 2010

Qy       1291 GCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGCATGCTGACAGCAGCCGCAA 1350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2011 GCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGCATGCTGACAGCAGCCGCAA 2070

Qy       1351 GCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGG 1410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2071 GCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGCTTTGGGTGATTGCCAAGAGCGG 2130

Qy       1411 CATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGT 1470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2131 CATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGT 2190

Qy       1471 CAGGCATGCGAGTAACAAGGGCATGCAACATCTGCTCAACATGAAGTGCAAAAATGTCGT 1530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2191 CAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGT 2250

Qy       1531 CCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACCGTCTTCATGGGGGTCGCAAGA 1590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2251 CCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACCGTCTTCATGGGGGTCGCAAGA 2310

Qy       1591 CTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCA 1650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2311 CTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCA 2370

Qy       1651 GAACCCACAGTCTCAGTGACGCCTGGCCCTGAGGTGAACTGGCCCACAGAGGTCACAAGC 1710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2371 GAACCCACAGTCTCAGTGACGCCTGGCCCTGAGGTGAACTGGCCCACAGAGGTCACAAGC 2430

Qy       1711 TGA 1713
              |||
Db       2431 TGA 2433


RESULT 5
AAD18442
ID   AAD18442 standard; cDNA; 2011 BP.
XX
AC   AAD18442;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human estrogen receptor (ER)-beta cDNA.
XX
KW   Estrogen receptor-beta; dietary compound; environmental pollutant;
KW   red wine; ER-beta mediated cell growth inhibition modulator; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers

FT   CDS             419..2011
FT                   /*tag= a
FT                   /product= "Human ER-beta protein"
XX
PN   WO200169262-A1.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-US08276.
XX
PR   15-MAR-2000; 2000US-189605P.
XX
PA   (UYMA-) UNIV MASSACHUSETTS.
XX
PI   Ho S;
XX
DR   WPI; 2001-596920/67.
DR   P-PSDB; AAE10457.
XX
PI   In vitro screening for modulator of estrogen receptor-beta-mediated
PT   cell growth inhibition by contacting cell having functional ER-beta
PT   protein with candidate compound and detecting ER-beta-regulated gene
PT   expression -
XX
PS   Example 3; Fig 7B; 49pp; English.
XX
CC   The invention relates to in vitro screening for modulator of estrogen
CC   receptor-beta (ER-beta)-mediated cell growth inhibition. The method
CC   involves contacting mammalian cell having functional ER-beta protein
CC   with candidate compound and detecting increase/decrease of ER-beta-
CC   regulated gene expression in presence of candidate compound compared
CC   to expression in absence of the compound. The method is useful for
CC   identifying a compound that modulates ER-beta-mediated cell growth
CC   inhibition. It is useful for testing dietary compounds, e.g. red wine,
CC   for the presence of estrogen mimics that trigger ER-beta-mediated
CC   processes, thus affecting physiological function positively or
CC   negatively and for testing environmental pollutants for the presence
CC   of estrogen mimics that may pose health risks involving ER-beta-
CC   mediated processes. The present sequence is human ER beta cDNA.
XX
SQ   Sequence 2011 BP; 470 A; 541 C; 561 G; 439 T; 0 other;

  Query Match             87.9%;  Score 1669;  DB 22;  Length 2011;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1669;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 402

Qy         61 GCTCTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 462

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 522

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 582

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 642

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 702

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 762
```

```
Qy        421 CACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCCCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        764 CACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCCCCAG 822

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        824 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 882

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 942

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 1002

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 1062

Qy        721 GAAGTGTGGCTCCCCGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 GAAGTGTGGCTCCCGGAGAGAGAGATCTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 1122

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 1182

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 1242

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1243 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 1302

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1303 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1362

Qy       1021 GATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGCAGAGCTCTTC 1422

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1482

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1542

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1602

Qy       1261 AGAATATCTCTGTGTCAACGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1603 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1662

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1722

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1782

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1842

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1902

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1962

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 2011
```

```
RESULT 6
AAS43105
ID   AAS43105 standard; cDNA; 2011 BP.
XX
AC   AAS43105;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human cDNA encoding oestrogen receptor beta.
XX
KW   Human; oestrogen receptor beta; ERbeta; ss; SNP; chromosome 6q.25.1;
KW   single nucleotide polymorphism; cardiovascular disease;
KW   autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;
KW   osteoarthritis; osteoporosis; breast cancer; endometrial cancer.
XX
OS   Homo sapiens.
XX
PN   WO200162793-A2.
XX
PD   30-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05360.
XX
PR   22-FEB-2000; 2000US-0183755.
PR   24-JAN-2001; 2001US-0768185.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Kalush F,  Cassel MJ,  Hwang SS,  Winn-Deen ES;
XX
DR   WPI; 2001-582041/65.
DR   P-PSDB; AA027322.
XX
PT   Oestrogen receptor gene and protein polymorphisms useful for diagnosis
PT   of individuals at risk of developing bone disorders -
XX
PS   Disclosure; Figure 3; 245pp; English.
XX
CC   The invention relates to a novel isolated peptide comprising or
CC   consisting of an amino acid sequence selected from an amino acid sequence
CC   of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
CC   10 amino acids), antibodies against them, nucleic acids encoding
CC   them (including vectors for transforming cells). The gene for human
CC   ERbeta is located on chromosome 6q.25.1. The variants are encoded
CC   by single nucleotide polymorphisms (SNP). The variant peptides and
CC   proteins can be used in assays to determine the biological
CC   activity of the protein, to raise antibodies, as a reagent in assays
CC   designed to quantitatively determine levels of the protein in
CC   biological fluids, to identify compounds that modulate receptor
CC   activity and to screen compounds for the ability to stimulate or
CC   inhibit interaction between the receptor protein and a target molecule
CC   that normally interacts with the receptor protein e.g. oestrogen.
CC   The antibody can be used to isolate the protein, to assess expression in
CC   disease states e.g. cardiovascular disease and autoimmune disease (e.g.
CC   systemic lupus erythematosus, arthritis, rheumatism and osteoarthritis),
CC   osteoporosis, breast cancer and endometrial cancer. In addition
CC   the antibodies can be used in pharmacogenomic analysis and inhibiting
CC   protein function, e.g. blocking the binding of the oestrogen receptor
CC   protein to a binding partner such as a ligand. The nucleic acids
CC   encoding the proteins can be used as probes, primers, chemical
```

CC intermediates and in biological assays. The present sequence encodes
CC human ERbeta.
XX
SQ Sequence 2011 BP; 470 A; 541 C; 561 G; 439 T; 0 other;

Query Match 87.9%; Score 1669; DB 22; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0.
Matches 1669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 402

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 462

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 522

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATACCCCTCCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 582

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTCCAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 TTACAGCATTCCCAGCAATGTCACTAACTTCCAAGGTGGGCCTGGTCGGCAGACCACAAG 642

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 702

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 762

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCCCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCCCCAG 822

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 882

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 942

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 1002

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACCAACTGCGAATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACCAACTGCGAATGCT 1062

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 1122

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 1182

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTGCTGCACGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTGCTGCACGC 1242

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1243 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATGAT 1302

Qy        961 CATCTCCCTGACCAAGTTGCCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1303 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1362

Qy       1021 GATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 GATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1422

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1482

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1542

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1602

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1662

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1722

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1782

Qy       1441 GGCTAACCTCGTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 GGCTAACCTCGTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1842

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1902

Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCGGAGTGCAGCCCGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCGGAGTGCAGCCCGGC 1962

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 2011
```

RESULT 7
AAX24364
ID AAX24364 standard; cDNA; 1686 BP.
XX
AC AAX24364;
XX
DT 07-JUN-1999 (first entry)
XX
DE Human oestrogen receptor-beta cDNA.
XX
KW Oestrogen receptor beta; hER beta; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 94..1686
FT /*tag= a
XX
PN WO9907847-A1.
XX
PD 18-FEB-1999.
XX
PF 20-JUL-1998; 98WO-US14944.
XX

```
PR   05 AUG 1997;   97US-0906365.
XX
PA   (AMHP ) AMERICAN HOME PROD CORP.
XX
PI   Bhat RA, Henderson RA, Hsiao C, Karathanasis SK;
XX
DR   WPI; 1999-167424/14.
DR   P-PSDB; AAW97858.
XX
PT   Human oestrogen receptor-beta coding sequence - useful in the
PT   production of human oestrogen receptor beta and identification of
PT   human oestrogen receptor-beta interactive compounds
XX
PS   Claim 1; Page 41-42; 49pp; English.
XX
CC   This cDNA clone encodes a full-length human oestrogen receptor-beta
CC   (hER-beta, see AAW97858) containing 45 previously unknown N-terminal
CC   amino acid residues that are believed to contribute to the
CC   transcription activation function of the receptor. Human testis
CC   cDNA was subjected to PCR using primers (see AAX24365-66) designed to
CC   supplement the previously known partial human sequence with 2
CC   missing N terminal residues (Met and Thr) found in rat ER-beta,
CC   and to supply an artificial Kozak translation initiation sequence.
CC   A fragment of approximately 1500 bp was obtained, and was cloned
CC   into vector pcDNA3 for sequencing. The 5' sequence was confirmed
CC   by PCR amplification from a human ovary cDNA library and by 5'RACE
CC   (see AAX24367-71). hER-beta is selectively expressed in the thymus,
CC   spleen, ovary and testes. The invention encompasses hER-beta
CC   polynucleotides and polypeptides, particularly peptides which
CC   include residues 1-45 of hER-beta. The invention also provides
CC   expression systems in which transcriptionally active hER-beta or
CC   fragments can be produced, as well as screening methods for
CC   identifying hER-beta agonists and antagonists (including
CC   tissue-specific oestrogens and anti oestrogens) and hER beta
CC   co-activators and inhibitors.
XX
SQ   Sequence 1686 BP; 413 A; 449 C; 441 G; 383 T; 0 other;

  Query Match             87.9%;  Score 1667.4;  DB 20;  Length 1686;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1668;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 77

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 GCTGCTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 137

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 197

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 257

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 317

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 377

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 437

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        438 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 497

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 557

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 617

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 677

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 737

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 797

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 857

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 917

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 977

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1037

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1097

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1157

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1217

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1277

Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1337

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1397

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1457

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1517

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1577
```

```
Qy       1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1637

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1686


RESULT 8
AAX34309
ID   AAX34309 standard; DNA; 1647 BP.
XX
AC   AAX34309;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Human oestrogen receptor coding sequence
XX
KW   Human; oestrogen receptor; ligand; bone resorption; metabolic disorder;
KW   cardiovascular disease; cancer; central nervous system; breast; uterine;
KW   osteoporosis; ovarian; prostate; diabetes; Alzheimer's disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9912961-A1.
XX
PD   18-MAR-1999.
XX
PF   04-SEP-1998;   98WO-US18577.
XX
PR   20-MAR-1998;   98GB-0006032.
PR   08-SEP-1997;   97US-0058271.
PR   30-SEP-1997;   97US-0060520.
PR   30-OCT-1997;   97GB-0022884.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Wilkinson H;
XX
DR   WPI; 1999-229222/19.
DR   P-PSDB; AAY07270.
XX
PT   Estrogen receptor
XX
PS   Claim 10; Fig 2; 32pp; English.
XX
CC   This is the coding sequence for a human oestrogen receptor.  The receptor
CC   can be used to identify ligands that bind to human oestrogen receptor.
CC   The ligands can be used in a method for preventing or treating an
CC   oestrogen receptor mediated disease or condition, such as abnormal bone
CC   resorption, a cardiovascular disease, cancer, metabolic disorders, or
CC   central nervous system disorders.  The ligand is especially used to treat
CC   osteoporosis, breast, uterine, ovarian or prostate cancer, diabetes or
CC   Alzheimer's disease.
XX
SQ   Sequence 1647 BP; 401 A; 438 C; 436 G; 372 T; 0 other;

  Query Match             86.2%;  Score 1635.6;  DB 20;  Length 1647;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1638;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAA 87
              | ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 CTTIGTAGCCICTTCTTGCAAGGTGTTTICTCAGCTGTTATCTCAAGACATGGATATAAA 65

Qy         88 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 125

Qy        148 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 185

Qy        208 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 245

Qy        268 CTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTCTGGCCAACACCTGG 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 CTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTCTGGCCAACACCTGG 305

Qy        328 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 365

Qy        388 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 425

Qy        448 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 485

Qy        508 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 545

Qy        568 GTCGTGTGAAGGATGTAAGGCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATAT 627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 GTCGTGTGAAGGATGTAAGGCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATAT 605

Qy        628 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTG 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTG 665

Qy        688 CCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCCTCCCGGAGAGAGAGATG 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 CCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCCTCCCGGAGAGAGAGATG 725

Qy        748 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 785

Qy        808 GGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAG 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 GGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAG 845

Qy        868 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 905

Qy        928 CCCCAGTGCCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 CCCCAGTGCCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 965

Qy        988 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1025

Qy       1048 GTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 GTTCGACCAAGTCCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1085

Qy       1108 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1145

Qy       1168 GGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 GGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1205

Qy       1228 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGAT 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGAT 1265
```

```
Qy       1288 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1325

Qy       1348 GAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1407
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 GAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1385

Qy       1408 CGGCATCTCCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1467
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 CGGCATCTCCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1445

Qy       1468 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1505

Qy       1528 GCCCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAA 1587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 GGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAA 1565

Qy       1588 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTC 1647
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1566 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTC 1625

Qy       1648 CCAGAACCCACAGTCTCAGTGA 1669
             |||||||||||||||||||||||
Db       1626 CCAGAACCCACAGTCTCAGTGA 1647


RESULT 9
AAT62843
ID   AAT62843 standard; cDNA; 1460 BP.
XX
AC   AAT62843;
XX
DT   08-JUN-1997  (first entry)
XX
DE   Human oestroqen receptor beta (ER-beta) cDNA.
XX
KW   Orphan receptor; oestroqen receptor beta; ER-beta;
KW   nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW   ovary cancer; cardiovascular disease; osteoporosis;
KW   environment; pollutant; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1460
FT                   /*tag= a
XX
PN   WO9709348-A2.
XX
PD   13-MAR-1997.
XX
PF   09-SEP-1996;   96WO-EP03933.
XX
PR   08-MAY-1996;   96GB-0009576.
PR   08-SEP-1995;   95GB-0018272.
PR   15-MAR-1996;   96GB-0005550.
PR   11-APR-1996;   96GB-0007542.
XX
PA   (KARO ) KARO BIO AB.
XX
PI   Enmark E, Gustafsson J, Kuiper GG;
XX
DR   WPI; 1997-192842/17.
DR   P-PSDB; AAW14724.
XX
PT   New isolated oestroqen receptor beta  used to develop prods. for
PT   treating e.q. cancers, CNS diseases, osteoporosis or cardiovascular
PT   disease
XX
PS   Claim 6; Fig 13B; 45pp; English.
```

```
XX
CC   A cDNA clone (AAI62843) codes for a novel human oestroqen receptor
CC   related nuclear receptor (AAW13724), designated ER-beta.  It was
CC   isolated from a human testis cDNA library using rat ER beta cDNA
CC   (see also AAT62842) as probe.  The human ER-beta gene was mapped
CC   to chromosome 14 reqion 14q22-23.  Rat, human and mouse ER-beta
CC   (AAW14723-25) can be used to isolate molecules for use in the
CC   treatment of cardiovascular diseases, central nervous system
CC   diseases, osteoporosis, prostate or ovarian cancer or benign
CC   prostatic hyperplasia and to test environmental chemicals for
CC   oestroqenic activity.
XX
SQ   Sequence 1460 BP; 352 A; 379 C; 411 G; 318 T; 0 other;

  Query Match             76.6%;  Score 1453.6;  DB 18;  Length 1460;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1456;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        210 CCATCACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 269
              | |||||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTATGACATTCTACAGTCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT  60

Qy        270 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 120

Qy        330 ACTTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 180

Qy        390 GTGCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 240

Qy        450 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 300

Qy        510 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 360

Qy        570 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 420

Qy        630 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 480

Qy        690 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTG 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTG 540

Qy        750 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 600

Qy        810 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 660

Qy        870 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 720

Qy        930 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 780

Qy        990 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 840
```

```
Qy       1050 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 1109
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 900

Qy       1110 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCCACATCTTCTTCTGCACAGGC 1169
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGC 960

Qy       1170 ATGAGGGCAAATGCGTAGAAGCAATTCTGGAAATCTTTGACATGCTGCTGGCAACTACTT 1229
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGAGGGCAAATGCGTAGAAGCAATTCTGGAAATCTTTGACATGCTGCTGGCAACTACTT 1020

Qy       1230 CAAGGTTTCGAGAGCTTAAAACTCCAACACAAAGAATATCTCTGTCTCAAGCCCATCATCC 1289
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGTTTCGAGAGCTTAAAACTCCAACACAAAGAATATCTCTGTCTCAAGCCCATCATCC 1080

Qy       1290 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1349
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1140

Qy       1350 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCC 1409
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCC 1200

Qy       1410 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACC 1469
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACC 1260

Qy       1470 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1529
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1320

Qy       1530 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1589
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1380

Qy       1590 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1649
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1440

Qy       1650 AGAACCCACAGTCTCAGTGA 1669
             |||||| ||||||||||||
Db       1441 AGAACCTACAGTCTCAGTGA 1460
```

```
RESULT 10
AAX26686
ID   AAX26686 standard; DNA; 1460 BP.
XX
AC   AAX26686;
XX
DT   17-JUN-1999  (first entry)
XX
DE   DNA encoding an estrogen receptor beta.
XX
KW   differential estrogen receptor; ER-alpha; ER-beta; activation;
KW   AP1 site; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             424..1881
FT                   /*tag= a
XX
PN   WO9911760-A1.
XX
PD   11-MAR-1999.
XX
PF   31-AUG-1998;   98WO-US18030.
XX
PR   04-SEP-1997;   97US-0923708.
XX
PA   (REGC ) UNIV CALIFORNIA
XX
PI   Gustafsson J, Kuiper GGJM, Kushner PJ, Nilsson S;
PI   Paech K, Scanlan TS, Webb P;
XX
DR   WPI; 1999-205173/17.
DR   P-PSDB; AAY01597.
XX
PT   Screening for test compounds which inhibit or activate estrogen
PT   receptor beta (ER beta) mediated activation at a AP1 site - for use
PT   in therapy
XX
PS   Example 1; Page 44 45; 61pp; English.
XX
CC   The present sequence encodes an estrogen receptor beta. The specification
CC   describes a method for screening test compounds for differential
CC   estrogen receptor (ER) alpha mediated and ER beta-mediated activation
CC   at an AP1 site. The method is used for identifying compounds
CC   which can be used in therapy.
XX
SQ   Sequence 1460 BP; 352 A; 379 C; 413 G; 316 T; 0 other;
```

```
  Query Match             76.4%;  Score 1450.4;  DB 20;  Length 1460;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1454;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        210 CCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 269
              | |||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db          1 CTATGACATTCTACAGTCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 60

Qy        270 TGGAAGCTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 120

Qy        330 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 180

Qy        390 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 240

Qy        450 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 300

Qy        510 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 360

Qy        570 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 629
              |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        361 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCAGGCAAGGACATAATGATTATATTT 420

Qy        630 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 480

Qy        690 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGATGTG 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGATGTG 540

Qy        750 CGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 600

Qy        810 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 660

Qy        870 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 720
```

```
Qy        940 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 780

Qy        990 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 840

Qy       1050 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 1109
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 900

Qy       1110 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 1169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 960

Qy       1170 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTT 1229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTT 1020

Qy       1230 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1289
              ||||||||||||||||||||||||||||||.||||||||||||||||||||||||||||
Db       1021 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1080

Qy       1290 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1140

Qy       1350 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1200

Qy       1410 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1260

Qy       1470 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1320

Qy       1530 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1380

Qy       1590 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1440

Qy       1650 AGAACCCACAGTCTCAGTGA 1669
              |||||| ||||||||||||
Db       1441 AGAACCTACAGTCTCAGTGA 1460


RESULT 11
AAT88412
ID   AAT88412 standard; cDNA; 1434 BP.
XX
AC   AAT88412;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human oestrogen receptor protein cDNA.
XX
KW   oestrogen receptor protein; steroid; alternative splicing; estradiol;
KW   estone; estriol; screening; ds.
XX
OS   Homo sapiens.
XX
PN   EP798378-A2.
XX
PD   01-OCT-1997.
XX
PF   25-MAR-1997;  97EP-0200903.
XX
PR   22-NOV-1996;  96EP-0203284.
PR   26-MAR-1996;  96EP-0200820.
XX
PA   (AKZO ) AKZO NOBEL NV.
XX
PI   Dijkema R,  Mosselman S;
XX
DR   WPI; 1997-473188/44.
DR   P-PSDB; AAW83212.
XX
PT   DNA encoding estrogen receptor - useful in screening assay to
PT   identify novel ligands or hormonal analogues
XX
PS   Claim 5; Page 13-14; 45pp; English.
XX
CC   This sequence encodes a novel oestrogen binding protein. The cDNA
CC   sequence which encodes this protein can be alternatively spliced
CC   resulting in the detection of additional transcripts (see AAT88413).
CC   This receptor is able to bind and be activated by estradiol, estone and
CC   estriol, can be used in a screening assay for the identification of new
CC   drugs e.g. novel ligands or hormonal analogues.
XX
SQ   Sequence 1434 BP; 347 A; 373 C; 406 G; 308 T; 0 other;

  Query Match             75.6%;  Score 1434;  DB 18;  Length 1434;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGGCCTGGTCGGCAGACC 295
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGGCCTGGTCGGCAGACC 60

Qy         296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy         356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy         416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy         476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy         536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy         596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy         656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy         716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy         776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy         836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
```

```
Qy        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAACATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCCGCCTCTTGCAGAGC 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGCAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCCGCAAC 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAAGTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAAGTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGCATGGTGACAGCAGCCCCAAGCTGGCTCACTTGCTGAACGCGGTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCCCAAGCTGGCTCACTTGCTGAACGCGGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCGTCCCAGCAGCAATCCATG 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCGTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260

Qy       1496 GAACATCTGCTCAACATGAAGTGCAAAAATGTGCTCCCAGTGTATGACCTGCTGCTGGAG 1555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGCTCCCAGTGTATGACCTGCTGCTGGAG 1320

Qy       1556 ATGCTGAATGCCCACCGTCTTCACGGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATGCTGAATGCCCACCGTCTTCACGGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380

Qy       1616 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTCA 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
```

```
RESULT 12
AAI72128
ID   AAI72128 standard; cDNA; 1434 BP.
XX
AC   AAI72128;
XX
DT   25-MAR-2002  (first entry)
XX
DE   Estrogen receptor cDNA.
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER-alpha; ER-beta; exon 8;
KW   estradiol; nuclear receptor; progesterone receptor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1434
FT                   /*tag= a
FT                   /product= "ER"
XX
PN   EP1162264-A2.
XX
PD   12 DEC-2001.
XX
PF   25-MAR-1997; 2001EP 9202021.
XX
PR   26-MAR-1996;   96EP-0200820.
PR   22 NOV 1996;   96EP 0203284.
PR   25-MAR-1997;   97EP-0200903.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Mosselman S,  Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   P-PSDB; AAB47834.
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor -
XX
PS   Example A; Page 12; 35pp; English.
XX
CC   This sequence encodes a novel estrogen receptor (ER). The gene encoding
CC   this new ER is located on chromosome 14 and has a different tissue
CC   distribution from classical ER. This ER also has two orphan ER's,
CC   ER-alpha and ER-beta. These orphan receptors have estrogen receptor
CC   related structure but do not appear to able to bind estradiol or other
CC   ER ligands. The DNA binding domain (DBD) and ligand binding domain (LBD)
CC   from this ER may be used in the chimeric receptor of the invention
CC   which also has an N-terminal domain. The chimeric receptor, or DNA
CC   encoding it, is useful in a screening assay for identification of new
CC   drugs. Similar chimeric receptors comprising the LBD of the new ER,
CC   and also comprising the DBD and an N-terminal domain derived from
CC   another nuclear receptor e.g., progesterone receptor, are useful for the
CC   screening of compounds to identify new ligands or hormone analogs which
CC   are able to activate the new ER. Chimeric receptors comprising a DBD
CC   of the new ER, and LBD and an N-terminal domain derived from
CC   another nuclear receptor, can be used to identify new ligands or
CC   hormone analogs for the nuclear receptors.
XX
SQ   Sequence 1434 BP; 347 A; 373 C; 406 G; 308 T; 0 other;

  Query Match             75.6%;  Score 1434;  DB 24;  Length 1434;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
```

```
Qy        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        716 ATGGTGAAGTGTGGCTCCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260

Qy       1496 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320

Qy       1556 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380

Qy       1616 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
```

```
RESULT 14
AAZ89597
ID   AAZ89597 standard; cDNA; 1820 BP.
XX
AC   AAZ89597;
XX
DT   23-JUN-2000  (first entry)
XX
DE   Bovine ER-beta cDNA.
XX
KW   Bovine; ER-beta; estrogen receptor beta; anticancer anti osteoporosis;
KW   transcriptional transactivator; estrogen-dependent transcription; cancer;
KW   estrogen-response element; ovary; prostate; osteoporosis; ds.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             153..1577
FT                   /*tag= a
FT                   /product= "ER-beta"
XX
PN   DE19839115 A1.
XX
PD   09-MAR-2000.
XX
PF   27-AUG-1998;   98DE-1039115.
XX
PR   27-AUG-1998;   98DE 1039115.
XX
PA   (IBFH-) IBF INST HORMON & FORTPFLANZUNGS.
XX
PI   Walther N;
XX
DR   WPI; 2000-257690/23.
DR   P-PSDB; AAY51946.
XX
PT   New bovine estrogen-beta receptor, useful for identifying specific
PT   ligands, potentially used for treating cancer and osteoporosis
XX
PS   Claim 4; Page 6-8; 20pp; German.
XX
CC   This invention describes a novel isolated bovine estrogen receptor-beta
CC   (ERb) which has anticancer and anti-osteoporosis activity. ERb is a
CC   transcriptional transactivator that activates estrogen-dependent
CC   transcription of genes that include an estrogen-response element. It is
CC   significantly less active than estrogen receptor-alpha. ERb (also its
CC   isoforms and peptides containing its ligand-binding domain), nucleic acid
CC   (III) encoding them, or vectors and cells containing (III) are used to
CC   identify ligands (agonists or antagonists) of ERb, particularly specific
CC   ligands that do not bind to the estrogen receptor alpha. These ligands
CC   are potentially useful for treating ERb-associated disorders,
CC   particularly cancer of the ovary or prostate, or osteoporosis. This
CC   sequence encodes the bovine ER-beta protein described in the method of
CC   the invention.
XX
SQ   Sequence 1820 BP; 446 A; 491 C; 478 G; 405 T; 0 other;

  Query Match             66.2%;  Score 1256.4;  DB 21;  Length 1820;
  Best Local Similarity   86.1%;  Pred. No. 0;
  Matches 1419;  Conservative    0;  Mismatches  216;  Indels   13;  Gaps    2;

Qy         84 TAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCT 143
              | ||||||||||||||||||||| |||||| |||||| |||||| ||||||||||
Db          1 TCAAAAACTCACCATCTAGCCTTAACTCTCCTGTCTCCTATAACTGCGGTCAATCCATCC 60

Qy        144 TACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAAT 203
              |||||||||||  ||| |||| || |||| || |||||||||| |||| ||||||||
Db         61 TACCCCTGGAGCCAGGCCCCATCTATTTACCCTCTTCCTATGTAGAGAGCCGCCATGAAT 120

Qy        204 ATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCA 263
              |  ||||| |||||||||||||||||||||||||||||||||| |||||         |
Db        121 ACTCAGCCGTGACATTCTATAGCCCTGCTGTGATGAATTACAGTATTCC    ---  -CG 171
```

```
Qy        264 CTAACTTGGAAGGTGGGCCTGCTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACAC 323
              ||||| |||| ||||||||||| |||||||||| |||||||| ||||||||||| |
Db        172 ATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGTCCAAATGTATTGTGGCCAACTC 231

Qy        324 CTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTC 383
              |||| ||||| |||||||||| ||||| ||||  ||| ||||||||| |||||||
Db        232 CTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCGGTCAGTTCTGTATGCAGAACCTC 291

Qy        384 AAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGA 443
              |||||||||| ||| ||||| | |||||  ||||||||||||||||||| ||||| ||||
Db        292 AAAAGAGTCCTTGGCGTGAAACCAGATCATTAGAACACACCTTACCTGTGAACAGGGAGA 351

Qy        444 CACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAA 503
              |||||||||||||| ||||||| | |||| |||||||||| |||| ||||| ||||||
Db        352 CACTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGCTACTAGTCCAAGTTCAA 411

Qy        504 AGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAG 563
              |||||||||| |||||||| |||||||||| ||||| || |||||||||||||||||||
Db        412 AGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCGGGATATCACTATCCAC 471

Qy        564 TCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATT 623
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        472 TCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGTATTCAAGGACATAATGATT 531

Qy        624 ATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGG 683
              |||||||||||||||||||||||||| || ||||| ||||| || || ||||| ||||
Db        532 ATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGACGTAAAAGCTGTCAGG 591

Qy        684 CCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGACAGACA 743
              |||| |||||||||||||| || || || ||||||||||||||||||| |||||||| |
Db        592 CCTGTCGACTTCGGAAGTGCTATGAGGTTGGAATGGTGAAGTGTGGCTCTCGGACACAAA 651

Qy        744 GATGTGGGTACCGCCTTGTGCGGACACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCG 803
              | |||||||||||| | ||||| || ||||||| | | || ||||||||||||||
Db        652 GGTGTGGGTACCGCATAGTGCGAAGGCAGAGAAATTCTGATGAGCAGCTGCCACTCCCTCA 711

Qy        804 GCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGACTGCCGCAGCTGCTGCTCCACCCCC 863
              |||| ||||||| |  || || | |  | |||||||   |||||||||||  |||||
Db        712 GCAAAACCAAGAGGAACGGTGGACCCATGACCGAGTGAAAGAGCTGCTGCTCAACCCCC 771

Qy        864 TGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCA 923
              ||||||| ||||||||| |||||  || |||||||||| |||||||||| |||||  || |
Db        772 TGAGCCCAGAGCAGCTGGTGCTTACGCTCCTGGAGGCCGAGCCGCCCCACGTGCTCATAA 831

Qy        924 GCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCG 983
              |||||||||  |||||||||| ||||||||||||||||||||||| |||||| ||||||
Db        832 GCCGCCCCAGCACGCCCTTCACTGAGGCCTCCATGATGATGTCCCTCACCAAGCTGGCCG 891

Qy        984 ACAAGGAGCTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGCAGCTCA 1043
              |||||||  |||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db        892 ACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCGGGCTTCGTGGAGCTCA 951

Qy       1044 GCCTGTTCGACCAAGTGCGGCTCTTTGGAGAGCTGCTTGGATGGAGGTGTTAATGATGGGC 1103
              |||||| |||||||||||||| |||||||||| ||| |||||||| | ||| ||||||
Db        952 GCCTGTACGACCAAGTGCGGCTTTTGGAGAGCTGCTGGTTGGAGGTGCTCATGGTGGGGC 1011

Qy       1104 TGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTCTTCTGC 1163
              |||||||||||| || |||||||| |||||||||||||||||||||| ||  ||||||
Db       1012 TGATGTGGCGCTCCATCGACCACCCTGGCAAGCTCATCTTTGCTCCAGACCTCATTCTGC 1071

Qy       1164 ACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTGCTGGCAA 1223
              |||||||||| |||||||| || |||||||||| ||||||||||||||||||||||||||
Db       1072 ACAGGGATGAAGGGAAATGTGTTGAAGGAATTCTAGAAATCTTTGACATGCTCCTGGCAA 1131

Qy       1224 CTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCA 1283
              | |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1132 CGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCA 1191

Qy       1284 TGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCA 1343
              ||||||| |||| ||||||||||||||   |  |||||  ||||||| ||||||||
Db       1192 TGATCCTCCTCAACTCCAGTATGTACCCTTCAGCTACAGCACCCCAGGAGGCTGACAGTG 1251

Qy       1344 GCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCCATGCTTTGCTTTGCGTGATTGCCA 1403
                  |||||||||| |||| ||||||| || |||| || ||| |||| ||||||||||
Db       1252 GCCGGAAGCTGACTCACCTGCTGAATGCTGTGACGGACGCTCTGGTCTGGGTGATTGCCA 1311

Qy       1404 AGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGT 1463
              |||| ||||| ||||||||||||| ||||||||||||||||||| ||||||||||||
Db       1312 AGAGTGGCATGTCCTCCCAGCAGCAGTCCATGCGCCTGGCTAACCTGCTGATGCTCCTGT 1371

Qy       1464 CCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAA 1523
              | |||||||||| || |||||||||||||||||||||||||| ||||||||||||||||||
Db       1372 CTCACGTCAGGCAGCGCAGTAACAAGGGCATGGAACACCTGCTCAACATGAAGTGCAAAA 1431

Qy       1524 ATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACAGTGCTTCGCGGGT 1583
              | |||||||| || || || ||||||||||||||||||||||||||||   ||||||||
Db       1432 ACGTGGTCCCCGTCTACGATCTGCTGCTGGAGATGCTGAATGCCCACACACTTCGCGGCA 1491

Qy       1584 GCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGG 1643
              |||||| |   |||| |||| ||| ||| || ||  |||||||| | |||||||||||
Db       1492 ACAAGTCTTTGGTCACAGGGTCTGAGCGCAACCTGGTGGAGGACAGTGAGAGCAAAGAGG 1551

Qy       1644 GCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCTGAGGTG----AACTGGCCCACAG 1699
              |||||||||| || ||| |||||||||||  |||| ||| ||    |   |
Db       1552 GCTCCCAGAAACCCCAGGCTCAGTGAGGCCCAGCCCCGAGCTGTGGTCAGACTGCAGAGA 1611

Qy       1700 AGGTCACAAGCTGAAGCGTGAACTCCAG 1727
              |||||| | |||||||| ||||||||||
Db       1612 AGGTCAGAGGCTGAAGCATGAACTCCAG 1639


RESULT 14
AAT88413
ID   AAT88413 standard; cDNA; 1251 BP.
XX
AC   AAT88413;
XX
DT   29-APR-1998  (first entry)
XX
DE   Human oestrogen receptor protein exon 8 splice variant cDNA.
XX
KW   oestrogen receptor protein, steroid, alternative splicing; estradiol;
KW   estone; estriol; screening; ds.
XX
OS   Homo sapiens.
XX
PN   EP798378-A2.
XX
PD   01-OCT-1997.
XX
PF   25-MAR-1997;   97EP-0200903.
XX
PR   22-NOV-1996;   96EP-0203284.
PP   26-MAR-1996;   96EP-0200820.
XX
PA   (AKZO ) AKZO NOBEL NV.
XX
PI   Dijkema R,  Mosselman S;
XX
DR   WPI; 1997-474188/44.
DR   P-PSDB; AAW33213.
XX
PT   DNA encoding estrogen receptor  useful in screening assay to
PT   identify novel ligands or hormonal analogues
XX
PS   Claim 5; Page 14-15; 45pp; English.
XX
CC   This partial sequence encodes a splice variant of a novel oestrogen
CC   binding protein, detected by screening a human testis cDNA library.
CC   This protein contains an alternative exon 8 (exon 8B) of the novel
CC   oestrogen receptor represented in AAT88412. As a consequence of the
CC   introduction of this exon through an alternative splicing reaction,
CC   the reading frame encoding the novel receptor is immediately terminated,
CC   creating a truncation of the carboxy terminus of the novel receptor.
CC   This novel receptor is able to bind and be activated by estradiol,
CC   estone and estriol, and can be used in a screening assay for the
```

```
CC   identification of new drugs e.g. novel ligands or hormonal analogues.
CC   however this variant does not contain an AF-2 region and therefore
CC   probably lacks the ability to modulate transcription of target genes in a
CC   ligand dependent fashion.
XX
SQ   Sequence 1251 BP; 297 A; 324 C; 354 G; 276 T; 0 other;

  Query Match             65.7%;  Score 1247;  DB 18;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        416 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        476 GCCAGCCCTGTTACTGGTCCAGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        776 AGTGCCGACGAGCAGCTGCACTGTGCCCGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 15
AAI72153
ID   AAI72153 standard; cDNA; 1251 BP.
XX
AC   AAI72153;
XX
DT   25-MAR-2002  (first entry)
XX
DE   cDNA encoding ER splice variant ER-beta.
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER-alpha; ER-beta; exon 8; PCR;
KW   estradiol; nuclear receptor; progesterone receptor; amplify; primer;
KW   polymerase chain reaction; AF-2 region; splice variant; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1251
FT                   /*tag=  a
FT                   /product= "ER-beta"
XX
PN   EP1162264-A2.
XX
PD   12-DEC-2001.
XX
PF   25-MAR-1997; 2001EP-0202021.
XX
PR   26-MAR-1996;   96EP-0200820.
PR   22-NOV-1996;   96EP-0203284.
PR   25-MAR-1997;   97EP-0200903.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Mosselman S,  Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   P-PSDB; AAB47837.
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor
XX
PS   Example B; Page 13; 35pp; English.
XX
CC   The sequences given in AAI72146 and AAI72153 encode splice variants
CC   of a novel estrogen receptor (ER). The gene encoding this new ER is
CC   located on chromosome 14 and has a different tissue distribution from
CC   classical ER.  This ER also has two orphan ER's, ER-alpha and ER-beta.
CC   These orphan receptors have estrogen receptor related structure but do
```

```
CC   not appear to able to bind estradiol or other ER ligands. These
CC   splice variants differ from the parent ER around exon 8. One clone
CC   contains exon 8B through alternative splicing. This causes an immediate
CC   termination of the reading thereby creating a truncation at the carboxy
CC   terminus. The other splice variant contains an alternative exon 8, exon
CC   8C, which encodes two C-terminal amino acids and then has a stop codon.
CC   The splice variant proteins do not contain the AF-2 region and therefore
CC   probably lack the ability to modulate transcription of target genes in
CC   a ligand dependant manner. However they may be able to interfere with
CC   the functioning of the WT classical ER or the novel ER of the invention,
CC   either by heterodimerisation or by occupying estrogen response elements
CC   or by interactions with other transcription factors.
XX
SQ   Sequence 1251 BP; 297 A; 324 C; 354 G; 276 T; 0 other;

  Query Match             65.7%;  Score 1247;  DB 24;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        356 CAGTTATCACATCTGTATGCGGAAGCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAAGCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCTGCGACTTCGGAAGTGTTACGAAGTGGGA 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCTGCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTCTGCCCAGACAGACA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTCTGCCCAGACAGACA 540

Qy        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTGCTG 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTGCTG 660

Qy        896 GAGGCTGAGCCCGCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCCGCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy       1076 TGTTGCATGGAGGTCTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 TGTTGCATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAGGGGAAATGCGTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGACTTAAAACTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGACTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 16
AAT88414
ID   AAT88414 standard; cDNA; 1257 BP.
XX
AC   AAT88414;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human oestrogen receptor protein splice variant cDNA (exon 8c).
XX
KW   Oestrogen receptor protein; steroid; alternative splicing; estradiol;
KW   estone; estriol; screening; ds.
XX
OS   Homo sapiens.
XX
PN   EP798378-A2.
XX
PD   01-OCT-1997.
XX
PF   25-MAR-1997;   97EP-0200903.
XX
PR   22-NOV-1996;   96EP-0203284.
PR   26-MAR-1996;   96EP-0200820.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Dijkema R, Mosselman S;
XX
DR   WPI; 1997-473188/44.
DR   P-PSDB; AAW33214.
XX
PT   DNA encoding estrogen receptor useful in screening assay to
PT   identify novel ligands or hormonal analoques
XX
PS   Claim 5; Page 28-29; 45pp; English.
XX
CC   This sequence encodes splice variant of a novel oestrogen binding
CC   protein isolated from human thymus tissue. This protein contain an
CC   alternative exon 8 (exon 8C) of the novel oestrogen receptor represented
CC   in AAT88412. This novel receptor is able to bind and be activated by
CC   estradiol, estone and estriol, can be used in a screening assay for the
CC   identification of new drugs e.g. novel ligands or hormonal analoques.
```

```
CC   This variant does not contain an AF 2 region and therefore probably lacks
CC   the ability to modulate transcription of target genes in a ligand
CC   dependent fashion.
XX
SQ   Sequence 1257 BP; 297 A; 427 C; 355 G; 278 T; 0 other;

  Query Match             65.7%;  Score 1247;  DB 18;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCC 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGGCGGCCACGCGCCC 600

Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        896 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 17
AA172146
ID   AA172146 standard; cDNA; 1257 BP.
XX
AC   AA172146;
XX
DT   25-MAR-2002  (first entry)
XX
DE   cDNA encoding ER splice variant ER-alpha.
XX
KW   DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW   estrogen receptor; ER; chromosome 14; ER-alpha; ER-beta; exon 8; PCR;
KW   estradiol; nuclear receptor; progesterone receptor; amplify; primer;
KW   polymerase chain reaction; AF-2 region; splice variant; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1257
FT                   /*tag= a
FT                   /product= "ER-alpha"
XX
PN   EP1162204-A2.
XX
PD   12-DEC-2001.
XX
PF   25-MAR-1997; 2001EP-0202021.
XX
PR   26-MAR-1996;   96EP-0200820.
PR   22-NOV-1996;   96EP-0203284.
PR   25-MAR-1997;   97EP-0200903.
XX
PA   (AKZO ) AKZO NOBEL NV.
XX
PI   Mosselman S,  Dijkema R;
XX
DR   WPI; 2002-084414/12.
DR   P-PSDB; AAB47836.
XX
PT   New isolated chimeric receptor comprising a DNA binding domain and/or
PT   ligand binding domain of a new estrogen receptor, for identifying
PT   functional ligands or hormonal analogs for the receptor
XX
PS   Example B; Page 24; 45pp; English.
XX
CC   The sequences given in AA172146 and AA172153 encode splice variants
CC   of a novel estrogen receptor (ER). The gene encoding this new ER is
CC   located on chromosome 14 and has a different tissue distribution from
CC   classical ER. This ER also has two orphan ER's, ER-alpha and ER beta.
CC   These orphan receptors have estrogen receptor related structure but do
CC   not appear to able to bind estradiol or other ER ligands. These
```

```
CC   splice variants differ from the parent ER around exon 8.  One clone
CC   contains exon 8B through alternative splicing.  This causes an immediate
CC   termination of the reading thereby creating a truncation at the carboxy
CC   terminus.  The other splice variant contains an alternative exon 8, exon
CC   8C, which encodes two C-terminal amino acids and then has a stop codon.
CC   The splice variant proteins do not contain the AF-2 region and therefore
CC   probably lack the ability to modulate transcription of target genes in
CC   a ligand dependant manner.  However they may be able to interfere with
CC   the functioning of the WT classical ER or the novel ER of the invention,
CC   either by heterodimerisation or by occupying estrogen response elements
CC   or by interactions with other transcription factors.
XX
SQ   Sequence 1257 BP; 297 A; 327 C; 355 G; 278 T; 0 other;

  Query Match             65.7%;  Score 1247;  DB 24;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATGGC 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATGGC 120

Qy        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        416 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGC 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGC 300

Qy        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        656 GATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        716 ATGGTGAACTGTGCCTCCGCGAGAGAGACATGTGCGTACCGCCTTGTGCGGAGACAGAGA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAACTGTGCCTCCGCGAGAGAGACATGTGCGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTC 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTC 660

Qy        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCC 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCTTCACCGAGGCCTCC 720

Qy        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGCAGTTGGTACACATGATCAGCTGGGCC 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGCAGTTGGTACACATGATCAGCTGGGCC 780

Qy       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy       1076 TGTTGGATGGAGGTGTTAATGATGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 TGTTGGATGGAGGTGTTAATGATGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 18
AAT62842
ID   AAT62842 standard; cDNA; 2568 BP.
XX
AC   AAT62842;
XX
DT   08-JUN-1997  (first entry)
XX
DE   Rat oestrogen receptor beta (ER-beta) cDNA.
XX
KW   Orphan receptor; oestrogen receptor beta; ER-beta;
KW   nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW   ovary cancer; cardiovascular disease; osteoporosis;
KW   environment; pollutant; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             424..1881
FT                   /*tag= a
FT                   /note= "alternative start codon at 448..450"
XX
PN   WO9709348-A2.
XX
PD   13-MAR-1997.
XX
PF   09-SEP-1996;   96WO-EP03933.
XX
PR   08-MAY-1996;   96GB-0009576.
PR   08-SEP-1995;   95GB-0018272.
PR   15-MAR-1996;   96GB-0005550.
PR   11-APR-1996;   96GB-0007532.
XX
PA   (KARO-) KARO BIO AB.
XX
PI   Enmark E;  Gustafsson J;  Kuiper GG;
XX
DR   WPI; 1997-192842/17.
DR   P-PSDB; AAW14723.
XX
PT   New isolated oestrogen receptor beta  used to develop prods. for
PT   treating e.g. cancers, CNS diseases, osteoporosis or cardiovascular
PT   disease
```

```
XX
PS   Claim 6; Fig 1; 45pp; English.
XX
CC   A cDNA clone (AAT62842) codes for a novel rat oestrogen receptor-
CC   related nuclear receptor (AAW14723), designated ER-beta.  A set of
CC   degenerate primers based on conserved sequences of the DNA-
CC   binding domain and ligand binding domain of nuclear receptors were
CC   used to amplify rat prostate cDNA.  The PCR prod. was used to
CC   screen a rat prostate cDNA library to obtain the ER-beta cDNA.
CC   This was then used to obtain human ER beta cDNA (see also AAT62843)
CC   from a human testis cDNA library.  Rat, human and mouse ER-beta
CC   (AAW14723 25) can be used to isolate molecules for use in the
CC   treatment of cardiovascular diseases, central nervous system
CC   diseases, osteoporosis, prostate or ovarian cancer or benign
CC   prostatic hyperplasia and to test environmental chemicals for
CC   oestrogenic activity.
XX
SQ   Sequence 2568 BP; 634 A; 654 C; 663 G; 617 T; 0 other;

  Query Match             65.3%;  Score 1239.6;  DB 18;  Length 2568;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1454;  Conservative    0;  Mismatches  279;  Indels    7;  Gaps    4;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              || || ||| ||||||||| ||      | | ||| |||||||  ||||| |||||||||
Db        213 CATGAGTCTCTGAGAACATAAT--GTCCATCTGTACCTCTTCTCACAAGGAGTTTTCTCA 270

Qy         61 GCTGTTA--TCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTC 117
               ||||  |   ||| |||||||||| || ||||||||||| || ||||||| ||| ||| |
Db        271 GCTGCGACCCTCTGAAGACATGGAGATCAAAAACTCACCGTCGAGCCTTAGTTC-CCTGC 329

Qy        118 CTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTC 177
              |||||  |||||| || || |||||| ||||||||||||||||| |||| ||||| |||||
Db        330 TTCCTATAACTGTAGCCAGTCCATCCTACCCCTGGAGCACGGCCCCATCTACATCCCTTC 389

Qy        178 CTCCTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGAT 237
              |||||| ||||||| || |||||| ||| |||| ||||||||||| || ||||||||||
Db        390 CTCCTACGTAGACAACCGCCATGAGTATTCAGCTATGACATTCTACAGTCCTGCTGTGAT 449

Qy        238 GAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCAC 297
              ||| |||||  ||||| ||||    || |||| |||| |||||||||  || | || |||
Db        450 GAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGCAC 509

Qy        298 AAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCA 357
              |||||||||||| | |||||||  |||||||||| |  |||| ||||
Db        510 AAGCCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGCCA 569

Qy        358 GTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGA 417
              | ||| |  || ||||| |||||||||||||||| ||||||||||||||| |||||
Db        570 ATCATCGCTGCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTAGA 629

Qy        418 ACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGC 477
              |||||||||||||||||||||||||||||||||  ||||| |||||||| | |||| ||
Db        630 GCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGTGC 689

Qy        478 CAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCCTGTCTGCAGCGA 537
              |||||||||||| |||||    |||||||||||||||||||| | ||||||||||
Db        690 CAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGCGA 749

Qy        538 TTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAA 597
              ||| ||||| || ||||| || || || ||||| ||||||||||||||||||||||||||
Db        750 TTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTTAA 809

Qy        598 AAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGA 657
              |||||||||||||||||||||||||||||||| |||||||| || |||||||||| || ||
Db        810 AAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATAGA 869

Qy        658 TAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAAT 717
              || |||||||| || |||||||||||||||||||||||| |||||||| ||||| |||||
Db        870 CAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGAAT 929

Qy        718 GGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAG 777
              ||| |||||||| ||| |||||||  | ||||||||||  | |||||||| ||||||||
Db        930 GGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGAAG 989

Qy        778 TGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCG 837
              | |   ||||||| | |||||      |||| ||||||||||  || || || || |||||
Db        990 TTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCCG 1049

Qy        838 AGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGA 897
              |||  |||||| ||||||  | || |||| || |||| || ||||||||||||||||||
Db       1050 GGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTGGA 1109

Qy        898 GGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCAT 957
              ||||| || ||| |||||||| | |||| ||||   ||||||||||||||||||||||
Db       1110 AGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCCAT 1169

Qy        958 GATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAA 1017
              |||||||||||| || ||| |||| |||||||  |||| ||||||||  |||||||||
Db       1170 GATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCCAA 1229

Qy        1018 GAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTG 1077
               ||| || || ||||||||||||||||||||||||| ||||||| ||||||| || |||||
Db        1230 GAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGCTG 1289

Qy        1078 TTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCT 1137
               |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||||
Db        1290 CTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAGCT 1349

Qy        1138 CATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCT 1197
               ||| || ||||| || || ||||||||||||||||||||||| |||||||||| |||||
Db        1350 CATTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGGATTCT 1409

Qy        1198 GGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTCGAGAGTTAAAACTCCAACA 1257
               |||||||||||||||||||||||| || || |||||||| || ||||||||||||| ||
Db        1410 GGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAGCA 1469

Qy        1258 CAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGT 1317
               ||| || |||||||||| ||||||||||||| |||| |||||||||||||  |||
Db        1470 CAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTGGC 1529

Qy        1318 CACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGAC 1377
               | || | |||||| || ||| | ||||||||| | ||| | |||||||| |||||
Db        1530 TTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTGAC 1589

Qy        1378 CGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCG 1437
               |||||  |||| |||||||||||| ||||| || ||||||||||||||||  ||  | ||
Db        1590 AGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCG 1649

Qy        1438 CCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGA 1497
               |||||  |||||||||||||| || || ||||||||||    ||||||||||||||||
Db        1650 ACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATGGA 1709

Qy        1498 ACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGAT 1557
               |||||||||||  |||||||||||||||||||||||||||  ||||||||||||||||||||
Db        1710 ACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGAT 1769

Qy        1558 GCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCC 1617
               |||||||| ||  |||||| |||| |||||||| ||| |||||||  |||||||| |
Db        1770 GCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGCTC 1829

Qy        1618 GGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGC 1677
               |||||||||||| | ||||||| |||||||||||  | |  |
Db        1830 AACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGATGGCCAGG 1889

Qy        1678 CCTGAGGTGAACTGGC-CCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGG 1736
               ||||||| | || | |  || ||| |    ||  || | | || || | ||  | |  ||
Db        1890 CCTGAGGCGGACAGACTACAGAGATGGTCAAAAGTGGAACATGTACCCTAGCATCTGGGG 1949


RESULT 19
AAX26685
ID   AAX26685 standard; DNA; 2568 BP.
XX
AC   AAX26685;
XX
```

```
DT   17-JUN-1999  (first entry)
XX
DE   DNA encoding an estrogen receptor-beta.
XX
KW   differential estrogen receptor; ER-alpha; ER-beta; activation;
KW   AP1 site; therapy; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             424..1881
FT                   /*tag= a
XX
PN   WO9911760-A1.
XX
PD   11-MAR-1999.
XX
PF   31-AUG-1998;   98WO-US18030.
XX
PR   04-SEP-1997;   97US-0923708.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Gustafsson J,  Kuiper GGJM,  Kushner PJ,  Nilsson S;
PI   Paech K,  Scanlan TS,  Webb P;
XX
DR   WPI; 1999-205173/17.
DR   P-PSDB; AAY01596.
XX
PT   Screening for test compounds which inhibit or activate estrogen
PT   receptor beta (ER beta) mediated activation at a AP1 site - for use
PT   in therapy
XX
PS   Example 1; Page 39-41; 61pp; English.
XX
CC   The present sequence encodes an estrogen receptor-beta. The specification
CC   describes a method for screening test compounds for differential
CC   estrogen receptor (ER) alpha-mediated and ER beta-mediated activation
CC   at an AP1 site. The method is used for identifying compounds
CC   which can be used in therapy.
XX
SQ   Sequence 2568 BP; 634 A; 652 C; 663 G; 619 T; 0 other;
```

Query Match 65.3%; Score 1239.6; DB 20; Length 2568;
Best Local Similarity 83.6%; Pred No. 0;
Matches 1454; Conservative 0; Mismatches 279; Indels 7; Gaps 4;

```
Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              || || ||| |||||||| ||     | | ||| ||||||||  ||||| |||||||||
Db        213 CATGAGTCTCTGAGAACATAAT--GTCCATCTGTACCTCTTCTCACAAGGAGTTTTCTCA 270

Qy         61 GCTGTTA---TCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTC 117
               ||||   |   ||| |||||||||| || |||||||||| || ||||||| ||| ||| |
Db        271 GCTGCGACCCTCTGAAGACATGGAGATCAAAAACTCACCGTCGAGCCTTAGTTC-CCTGC 329

Qy        118 CTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATAIACAIACCTTC 177
              ||||| ||||| || || |||||| |||||||||||||||| |||| ||||| |||||
Db        330 TTCCTATAACTGTAGCCAGTCCATCCTACCCCTGGAGCACGGCCCCAICIACATCCCTTC 389

Qy        178 CTCCTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGAT 237
              |||||| ||||||| || |||||| ||| |||| |||||||||| || ||||||||||
Db        390 CTCCTACGTAGACAACCGCCATGAGTATTCAGCTATGACATTCTACAGTCCTGCTGTGAT 449

Qy        238 GAATTACAGCATTCCCAGCAATGTCACTAACTTGCAAGGTGGGCCTGGTCGGCAGACCAC 297
              ||| |||||  ||||| |||    || |||| |||| |||||||||  || | || |||
Db        450 GAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGCAC 509

Qy        298 AAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCA 357
              |||||||||||||| | ||||||||  |||||||||| |||||||||| |  |||| ||||
Db        510 AAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGCCA 569

Qy        358 GTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGA 417
                | ||| |  || ||||| |||||||||||||||| |||||||||||||||| |||||

Db        570 ATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTAGA 629

Qy        418 ACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGC 477
              |||||||||||||||||||||||||||||||||||| |||||| |||||||| | |||| ||
Db        630 GCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGTGC 689

Qy        478 CAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTCGCGCTGTCTGCAGCGA 537
              |||||||||||||| |||||    |||||||||||||||||||||| | ||||||||||
Db        690 CAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTCGCCCGTCTGCAGCGA 749

Qy        538 TTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAA 597
              ||| ||||| || |||| || || || |||| ||||||||||||||||||||||||||||
Db        750 TTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTTAA 809

Qy        598 AAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGA 657
              |||||||||||||||||||||||||||||||| |||||||| || |||||||||| || ||
Db        810 AAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATAGA 869

Qy        658 TAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAAT 717
              || |||||||| || |||||||||||||||||||||||| |||||||| |||| |||||
Db        870 CAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGAAT 929

Qy        718 GGTGAAGTGTGGCTCCCGGAGACAGACATGTGGGTACCGCCTTGTGCCGAGACACAGAAG 777
              ||| |||||||| ||| ||||||| | ||||||||||  | |||||||| |||||||
Db        930 GGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGAAG 989

Qy        778 TGCCGACGAGCAGCTGCACTGTGCCGGCAAGCCCAAGAGAAGTGGCGGCCACGCGCCCCG 837
              | |   |||||| | ||||      |||| ||||||||||  || || || || |||||
Db        990 TTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCCCG 1049

Qy        838 AGTGGGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGA 897
              |||  |||||| |||||  | || |||| || |||| || ||||||||||||||||||
Db       1050 GGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTGGA 1109

Qy        898 GGCTGAGCCGCCCCATGTGCTGATCAGCGCCCCAGTGCGCCCTTCACCGAGGCCTCCAT 957
              ||||| || ||| |||||||| | ||||| |||||   ||||||||||||||||||||
Db       1110 AGCTGAACCACCCAATGTGCTGGGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCCAT 1169

Qy        958 GATGATGTCCCTCACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAA 1017
              |||||||||||| || ||| |||| |||||||   |||| |||||||   |||||||||
Db       1170 GATGATGTCCCTCACTAAGCTGGCGCACAAGAACTGGTGCACATGATTGGCTGGGCCAA 1229

Qy       1018 GAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCCGGCTCTTGGAGAGCTG 1077
              ||| || || |||||||||||||||||||||||| |||||||| |||||||| || |||||
Db       1230 CAAAATCCCTGGCTTTGTCGACCTCACCCTGTTCGACCAAGTCCGGCTCTTAGAAAGCTG 1289

Qy       1078 TTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCT 1137
              |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||||
Db       1290 CTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAGCT 1349

Qy       1138 CATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCT 1197
              ||| || |||| || || |||||||||||||||||||||||| ||||||||||| |||||
Db       1350 CATTTTCGCTCCCGACCTCGTTCTGCACAGGGATGAGGGGAAGTGCGTAGAAGGGATTCT 1409

Qy       1198 GGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACA 1257
              |||||||||||||||||||| || || |||||||| || ||||||||||||||| ||
Db       1410 GGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAGCA 1469

Qy       1258 CAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGT 1317
              ||| || |||||||||| ||||||||||||| |||| ||||||||||||  |||
Db       1470 CAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTGGC 1529

Qy       1318 CACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGAC 1377
                | || | |||||| || || || |||||||||||| | ||| | |||||||| |||||
Db       1530 TTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTGAC 1589

Qy       1378 CGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCG 1437
              |||||  |||| |||||||||| ||||| || ||||||||||||||||| ||  | ||
Db       1590 AGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCG 1649

Qy       1438 CCTGGCTAACCTCGCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGA 1497
              ||||| |||||||||||||| || || |||||||||||    ||||||||||||||||||
Db       1650 ACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATGGA 1709
```

```
Qy       1498 ACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGAT 1557
             ||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1710 ACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGAT 1769

Qy       1558 GCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCC 1617
             ||||||||| ||| |||||| |||| |||||||||| ||| ||||||| |||||||||  |
Db       1770 GCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGCTC 1829

Qy       1618 GGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCACTCACGCCTGCC 1677
                ||||||||||||  | ||||||| |||||||||||  |||||||||||| | |  |
Db       1830 AACAGAGGACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGATGGCCAGG 1889

Qy       1678 CCTGAGGTGAACTGGC CCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTCTCAGG 1736
              ||||||| | || | |  || ||| |    ||  || | | || || | ||  | |  ||
Db       1890 CCTGAGGCGGACAGACTACAGAGATGGTCAAAAGTGGAACATGTACCCTAGCATCTGGGG 1949


RESULT 20
AAX43382
ID   AAX43382 standard; DNA; 1704 BP.
XX
AC   AAX43382;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Murine mER beta-3 clone DNA sequence.
XX
KW   Estrogen receptor beta; oestrogen receptor beta; ER beta; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1680
FT                   /*tag= a
FT                   /product= "complete oestrogen receptor beta"
FT                   /note= "no stop codon given"
XX
PN   WO9905171-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;    98WO-US15540.
XX
PR   30-JUL-1997;    97US-0054210.
PR   28-JUL-1997;    97US-0053869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Baron R, Chambon P, Denton R, Dupont S, Garnier J;
XX
DR   WPI; 1999-142853/12.
DR   P-PSDB; AAY04433.
XX
PT   Novel murine oestrogen receptor beta genes and splice variants -
PT   useful for treating oestrogen receptor-beta mediated disorders
XX
PS   Example 1; Fig 1b 1; 89pp; English.
XX
CC   The present sequence encodes the murine complete oestrogen receptor beta
CC   (mER beta c; also designated mER beta 4). The protein can be used to
CC   screen for agents that modulate or block the interaction between the
CC   ER beta c protein and physiological, stilbene or triphenylethylene
CC   (anti)oestrogens. The antibody which binds the protein can be used in
CC   assays to detect activation of the ER beta protein and can be used for
CC   subcellular localisation of the protein. The complementary nucleic acid
CC   can be used to inhibit expression of the ER beta c nucleic acid.
XX
SQ   Sequence 1704 BP; 424 A; 457 C; 448 G; 375 T; 0 other;

  Query Match             61.1%;  Score 1160.2;  DB 20;  Length 1704;
  Best Local Similarity   81.8%;  Pred. No. 1.5e-296;
  Matches 1389;  Conservative    0;  Mismatches  253;  Indels   57;  Gaps    2;

Qy         28 CCTTTGTGGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCT--CAAGACATGGATAT 84
              | | ||||||||||||  |||||  ||||||||||||  | ||   |||||||||| ||
Db          6 CATCTGTGGCCTCTTCTCACAAGGATTTTTCTCAGCTGAGACCTACGCAAGACATGGAGAT 65

Qy         85 AAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTT 144
               |||||||||||||| |||||| ||| ||| | || || ||||| || || |||||| |
Db         66 CAAAAACTCACCATCAAGCCTTACTTCCCCTGCTTCTTATAACTGTAGCCAGTCCATCCT 125

Qy        145 ACCCCTGCAGCACGCCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATA 204
              |||| ||||||| ||  |||| || || ||||||||||||||||| ||||  || |||||
Db        126 ACCCTTGGAGCATGGTCCCATCTATATCCCTTCCTCCTATGTAGAGAGCCGTCACGAATA 185

Qy        205 TCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCAC 264
              |||||||||||||||| || |||||||||||||| |||||  |||||||||    |
Db        186 CTCAGCCATGACATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGG 245

Qy        265 TAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACC 324
              |||| |||||||||||||| ||| |||||  |||||||||||||| | ||||||||  |
Db        246 TAACCTGGAAGGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTGCTATGGCCAACTTC 305

Qy        325 TGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCA 384
              ||| |||||  ||||||||||    |||  |||| | ||| | ||| |||||  ||||||||
Db        306 TGGACACCTCTCTCCTTTAGCCACCCACTGCCAATCATCGCTTCTCTATGCAGAACCTCA 365

Qy        385 AAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGAC 444
              |||||||||  |||||||||||||||||  |||||||||||||  |||||||||||||||||
Db        366 AAAGAGTCCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGAC 425

Qy        445 ACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAA 504
              |||||  ||||||  || | |||| | |||| ||||||||||||||| |||| |  | ||
Db        426 CCTGAAGAGGAAGCTTGGCGGGAGCGGTTGTGCCAGCCCTGTTACTAGTCCAAGCGCCAA 485

Qy        505 GAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGT 564
              |||||||||||||||||||| |||||||| ||||| |||||  || |||||  || || ||
Db        486 GAGGGATGCTCACTTCTGCGCCGTCTGCAGTGATTATGCATCTGGGTATCATTACGGTGT 545

Qy        565 CTGGTCGTGTGAAGCGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTA 624
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        546 CTGGTCCTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTA 605

Qy        625 TATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGC 684
              ||| |||||||| || |||||||||| || || || |||||||| || ||||||||||
Db        606 TATCTGTCCAGCCACGAATCAGTGTACCATAGACAAGAACCGGCGTAAAAGCTGCCAGGC 665

Qy        685 CTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGGAGAGAGAG 744
              |||||||||||| |||||||||||||| |||||||| |||||||| ||| |||||| ||
Db        666 CTGCCGACTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGAGAAAG 725

Qy        745 ATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACCAGCAGCTGCACTGTGCCGG 804
              ||||||||||  | || || ||||||||||||||  |||||||| |||| ||
Db        726 GTGTGGGTACCGAATAGTACGAAGACAGAGAAGTGCCAGCGAGCAGGTGCATTGCCTGAA 785

Qy        805 CAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCT 864
              ||| ||||||||||  | || ||| | |||| ||| |||||| |||| || | ||
Db        786 CAAAGCCAAGAGAACCAGTGGGCACACACCCCGGGTGAAGGAGCTACTGCTGAACTCTCT 845

Qy        865 CACCCCCGAGCAGCTAGTCCTCACCCTCCTGCAGCCTCAGCCGCCCCCATGTCCTCATCAC 924
              ||| |||||||||| |||||||||| |||| |||||||| ||| |||||||  | ||
Db        846 GAGTCCCGAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTGCTAGTGAG 905

Qy        925 CCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGA 984
              ||| ||||  |||||||||||||||||||||||||||||||||||| || ||| |||| ||
Db        906 CCGTCCCAGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTCACGAAGCTGGCTGA 965

Qy        985 CAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAG 1044
              ||||||  |||| ||||||||  |||||||||||||| || || ||||||||||||||||
Db        966 CAAGGAACTGGTGCACATGATTGGCTGGGCAAGAAAATCCCTGGCTTTGTGGAGCTCAG 1025

Qy       1045 CCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCT 1104
              |||||| |||||||| || |||||||| ||||| ||||||||||| | ||| |||||||
Db       1026 CCTGTTGGACCAAGTCCGCCTCTTGGAAAGCTGCTGGATGGAGGTGCTGATGGTGGGGCT 1085
```

```
Qy       1105 GATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGA 1164
             ||||||||||| || ||||||||||||||||||||||||||||||||| || ||||||||
Db       1086 GATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGA 1145

Qy       1165 C------------------------------------------------------AGGGA 1170
              |                                                      |||||
Db       1146 CAGGTCCTCAGAAGACCCTCACTGGCACGTTGCGCAGACGAAGAGTGCTGTGCCAAGGGA 1205

Qy       1171 TGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTC 1230
             |||||||| ||||| ||||| ||||||||||||||||||||||||||||| || ||  |
Db       1206 TGAGGGGAAGTGCGTGGAAGGGATTCTGGAAATCTTTGACATGCTCCTGGCGACGACGGC 1265

Qy       1231 AAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCT 1290
             | |||| || |||||||||| || ||||||||||||| ||||| |||||||||| ||
Db       1266 ACGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCATGATTCT 1325

Qy       1291 GCTCAATTGCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAA 1350
              ||||| ||||||||||||||  |||  || || | |||||| || || || |||||||
Db       1326 CCTCAACTCCAGTATGTACCCCTTGGCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAA 1385

Qy       1351 GCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGG 1410
             |||| | ||| |  |||||| ||||| |||||  |||| |||||||| | ||||| ||
Db       1386 GCTGACACACCTATTGAAGGCAGTGACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTGG 1445

Qy       1411 CATCTCCTCCCAGCAGCAATGCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGT 1470
              ||||| |||||||||| ||  | || ||||| |||||||||||| || || || ||
Db       1446 AATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGT 1505

Qy       1471 CAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAAIGTGGI 1530
             ||||||    |||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1506 CAGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGI 1565

Qy       1531 CCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTC 1590
             ||| ||||| |||||||||||||||||||||||| ||| ||||| |||| ||||||
Db       1566 CCCGGTGTACGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAGTC 1625

Qy       1591 CTCCATCACGGGGTCCGAGTGCAGCCCGCCAGAGGACAGTAAAAGCAAAGAGGGCTCCCA 1650
             ||| ||| ||||||| |||||| || || |||||||||||| ||||||||||||||||||
Db       1626 CTCAATCTCGGGGTCTGAGTGCTGCTCGACAGAGGACAGTAAGAGCAAAGAGGGCTGCCA 1685

Qy       1651 GAACCCACAGTCTCAGTGA 1669
              |||||  ||||| ||||||
Db       1686 GAACCTCCAGTCACAGTGA 1704


RESULT 21
AAX24935
ID   AAX24935 standard; cDNA; 1704 BP.
XX
AC   AAX24935;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Mouse complete oestrogen receptor beta clone mERbeta-3.
XX
KW   oestrogen receptor; ER beta-3; ER beta-c; mouse; tumour; diagnosis;
KW   therapy; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   exon            1149..1203
FT                   /*tag= a
FT                   /number= 5B
XX
PN   WO9905170-A1.
XX
PD   04-FEB-1999.
XX
PF   28-JUL-1998;   98WO-US15539.
XX
PR   30-JUL-1997;   97US-0054210.
PR   28-JUL-1997;   97US-0053869.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Denton R,  Frasion C,  Galien R,  Rische-Rigon M;
XX
DR   WPI; 1999-142852/12.
DR   P-PSDB; AAW98128.
XX
PT   Human oestrogen receptor beta-3 - useful in treating and monitoring
PT   tumours
XX
PS   Example 1; Fig 1b; 89pp; English.
XX
CC   The present sequence is the coding region of the murine complete
CC   oestrogen receptor-beta (ER beta-c) cDNA clone mER beta-3, which
CC   encodes a 567 amino acid protein (see AAW98128).  The gene contains
CC   9 exons, including the newly described exon 5B, which is not found
CC   in ER beta-1.  Exon 5B is located in the ligand binding domain of
CC   mER beta-3, and probably plays a significant role in mER beta-3
CC   function.  The mER beta-3 clone was twice isolated using 2
CC   separate procedures: RT-PCR of mRNA; and amplification from a
CC   mouse embryonic stem cell genomic DNA library.  2 Alternatively
CC   spliced forms of mER beta-3 were also identified; mER beta-1 (see
CC   AAW98125), which lacks exon 5B; and mER beta-2 (see AAW98126), which
CC   lacks exon 3.  A third isoform, rER beta-4 (see AAW98127) was
CC   identified in rat ovaries.  The human exon 5B region (see AAX24934)
CC   has also been identified.  Based on these sequences, the invention
CC   provides methods for identifying agents that block or augment
CC   ER beta-c mediated transcriptional regulation, methods to determine
CC   whether ER beta c or its isoforms is being expressed in tissues or
CC   cells, and methods of identifying and using agents that block the
CC   transcriptional regulation of genes by ER beta-c or its isoforms,
CC   which in turn modulates other biological and pathological
CC   processes.  Gene therapy expression of ER beta-c is envisaged.
XX
SQ   Sequence 1704 BP; 424 A; 458 C; 447 G; 375 T; 0 other;

  Query Match             61.0%;  Score 1158.6;  DB 20;  Length 1704;
  Best Local Similarity   81.7%;  Pred. No. 3.9e-296;
  Matches 1388;  Conservative    0;  Mismatches  254;  Indels   57;  Gaps    2;

Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCT---CAAGACATGGATAT 84
              | | ||||||||||||  |||||  ||||||||||||  | ||   ||||||||||| ||
Db          6 CATCTGTGCCTCTTCTCACAAGGATTTTTCTCAGCTGAGACCTACGCAAGACATGGAGAT 65

Qy         85 AAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTT 144
              ||||||||||||||| ||||||| ||| ||| | || || |||| || || |||||| |
Db         66 CAAAAACTCACCATCAAGCCTTACTTCCCCTGCTTCTTATAACTGTAGCCAGTCCATCCT 125

Qy        145 ACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATA 204
              |||| ||||||| ||  |||| || || ||||||||||||||||| |||| || |||||
Db        126 ACCCTTGGAGCATGGTCCCATCTATATCCCTTCCTCCTATGTAGAGAGCCGTCACGAATA 185

Qy        205 TCCACCCATCACATTCTATAGCCCTGCTCTCATGAATTACAGCATTCCCACCAATCTCAC 264
              |||||||||||||||| || ||||||||||||| |||||  |||||||||      |
Db        186 CTCAGCCATCACATTCTACAGTCCTGCTCTCATGAACTACAGTCTTCCCAGCAGCACCGG 245

Qy        265 TAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACC 324
              |||| |||||||||||||| ||| |||||  |||||||||||| | ||||||||  |
Db        246 TAACCTGGAAGGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTGCTATGGCCAACTTC 305

Qy        325 TGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCA 384
              ||| ||||| |||||||||    ||| |||| | ||| | ||| ||||| ||||||||
Db        306 TGGACACCTCTCTCCTTTAGCCACCCACTGCCAATCATCGCTTCTCTATGCAGAACCTCA 365

Qy        385 AAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGAC 444
              |||||||| |||||||||||||||| |||||||||||||  ||||||||||||||||||
Db        366 AAAGAGTCCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGAC 425

Qy        445 ACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAA 504
              ||||| |||||| || | |||| | |||| ||||||||||||||  |||| |  | ||
Db        426 CCTGAAGAGGAAGCTTGGCGGCAGCGGTTGTGCCAGCCCTGTTACTACTCCAAGCGCCAA 485
```

```
Qy        505 GAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGT 564
              ||||||||||||||||||||| |||||||| ||||| ||||| || ||||| || || ||
Db        486 GAGGGATGCTCACTTCTGCGCCGTCTGCAGTCATTATGCATCTGGGTATCATTACGGTGT 545

Qy        565 CTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTA 624
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        546 CTGGTCCTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTA 605

Qy        625 TATTTGTCCAGCTACAAATCACGTGTACAATCGATAAAAACCCGCGCAAGAGCTGCCAGGC 684
              ||| |||||||| || |||||||||| || || || |||||||| || ||||||||||
Db        606 TATCTGTCCAGCCACGAATCAGTGTACCATACACAAGAACCGGCGTAAAAGCTGCCAGGC 665

Qy        685 CTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAG 744
              |||||||||||| |||||||||||||| |||||||| |||||||| ||| ||||||| ||
Db        666 CTGCCGACTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGAGAAAG 725

Qy        745 ATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGG 804
              |||||||||||  | || || |||||||||||||||  ||||||| |||| ||
Db        726 GTGTGGGTACCGAATAGTACGAAGACAGAGAAGTGCCAGCGAGCAGGTGCATTGCCTGAA 785

Qy        805 CAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCT 864
              ||| ||||||||||    | || ||| | ||||| |||  |||||| |||||| || | ||
Db        786 CAAAGCCAAGAGAACCAGTGGGCACACACCCCGGGTGAAGGAGCTACTGCTGAACTCTCT 845

Qy        865 GACCCCCGCAGCAGCTAGTGCTCACCCTCCTGGAGACTGAGCCGCCCCATGTGCTGATCAG 924
              ||| |||||||||| |||||||||| ||||| ||||||| ||| ||||||| | ||
Db        846 GAGTCCCGAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTGCTAGTGAG 905

Qy        925 CCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGCCCGA 984
              ||| |||||   |||||||||||||||||||||||||||||||| || ||| |||| ||
Db        906 CCGTCCCAGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTCACGAAGCTGGCTGA 965

Qy        985 CAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAG 1044
              |||||| |||| |||||||| |||||||||||| || || ||||||||||||||||||
Db        966 CAAGGAACTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAG 1025

Qy       1045 CCTGTTCGACCAAGTGCTGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGCT 1104
              |||||| |||||||| || |||||||| ||||| |||||||||||| | ||| |||||||
Db       1026 CCTGTTGGACCAAGTCCGCCTCTTGGAAAGCTGCTGGATGGAGGTGCTGATGGTGGGCT 1085

Qy       1105 GATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGA 1164
              |||||||||||| || ||||||||||||||||| |||||||||||| || ||||||||
Db       1086 GATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGA 1145

Qy       1165 C - - - - - - - - - - - - - - - - - - - - - - - - - - - AGGGA 1170
              |                                                     |||||
Db       1146 CAGGTCCTCAGAAGACCCTCACTGGCACGTTGCGCAGACGAAGAGTGCTGTCCCAAGGGA 1205

Qy       1171 TGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTC 1230
              |||||||| |||| |||| |||||||||||||||||||||||||||||| || ||  |
Db       1206 TGAGGGGAAGTGCGTGGAAGGCATTCTGGAAATCTTTGACATGCTCCTGGCGACGACGGC 1265

Qy       1231 AAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCT 1290
              | |||| || |||||||||| || |||||||||||||| ||||| |||||||||| ||
Db       1266 ACGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCATGATTCT 1325

Qy       1291 GCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAA 1350
              |||||  |||||||||||||||  ||| || || | ||||| || || || ||||||||
Db       1326 CCTCAACTCCAGTATCTACCCCTTGCCTACCCCAAGCCAGGAAGCAGAGAGTAGCCGGAA 1385

Qy       1351 GCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGG 1410
              |||| | ||| |  || |||| ||||| |||||  |||| ||||||||| | ||||| ||
Db       1386 GCTGACACACCTATTGAACGCAGTGACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTGG 1445

Qy       1411 CATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCACGT 1470
              ||||| |||||||||||| ||  | || ||||| |||||||||||||| || || || ||
Db       1446 AATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGT 1505

Qy       1471 CAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGT 1530
              ||||||    |||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1506 CAGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGT 1565

Qy       1531 CCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTC 1590
```

```
              ||| ||||| |||||||||||||||||||||||||| |||  |||||| |||| ||||||
Db       1566 CCCGGTGTACGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAGTC 1625

Qy       1591 CTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCA 1650
              ||| ||| |||||| |||||| || || |||||||||||| ||||||||||||||||||
Db       1626 CTCAATCTCGGGGTCTGAGTGCTGCTCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCA 1685

Qy       1651 GAACCCACAGTCTCAGTGA 1669
              |||||  ||||| ||||||
Db       1686 GAACCTCCAGTCACAGTGA 1704


RESULT 22
AAT62844
ID   AAT62844 standard; cDNA; 1458 BP.
XX
AC   AAT62844;
XX
DT   08-JUN-1997  (first entry)
XX
DE   Mouse oestrogen receptor beta (ER-beta) cDNA.
XX
KW   Orphan receptor; oestrogen receptor beta; ER beta;
KW   nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW   ovary cancer; cardiovascular disease; osteoporosis;
KW   environment; pollutant; ss.
XX
OS   Mus sp.
XX
PN   WO9709348-A2.
XX
PD   13-MAR-1997.
XX
PF   09-SEP-1996;   96WO-EP03934.
XX
PR   08-MAY-1996;   96GB-0009576.
PR   08-SEP-1995;   95GB-0018272.
PR   15-MAR-1996;   96GB-0005550.
PR   11-APR-1996;   96GB-0007532.
XX
PA   (KARO-) KARO BIO AB.
XX
PI   Enmark E, Gustafsson J, Kuiper GG;
XX
DR   WPI; 1997-192842/17.
DR   P-PSDB; AAW14725.
XX
PT   New isolated oestrogen receptor beta  used to develop prods. for
PT   treating e.g. cancers, CNS diseases, osteoporosis or cardiovascular
PT   disease
XX
PS   Claim 6; Fig 14B; 45pp; English.
XX
CC   A cDNA clone (AAT62844) codes for a novel mouse oestrogen receptor
CC   related nuclear receptor (AAW14725), designated ER beta.  Rat, human
CC   and mouse ER beta (AAW14723 25) can be used to isolate molecules for
CC   use in the treatment of cardiovascular diseases, central nervous
CC   system diseases, osteoporosis, prostate or ovarian cancer or benign
CC   prostatic hyperplasia and to test environmental chemicals for
CC   oestrogenic activity.
XX
SQ   Sequence 1458 BP; 365 A; 373 C; 403 G; 317 T; 0 other;

  Query Match             57.9%;  Score 1099.6;  DB 18;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 1.5e-280;
  Matches 1234;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy         212 ATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTG 271
               ||| ||||||| || |||||||||||||| |||||  |||||||||    | |||| ||
Db           1 ATGGCATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTG 60

Qy         272 GAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCAC 331
               |||||||||||| ||| ||||| |||||||||||||| | ||||||||  |||| |||
```

```
Db        61 GAAGGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTGCTATGGCCAACTTCTGGACAC 120

Qy       342 CTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGT 391

Db       121 CTCTCTCCTTTAGCCACCCACTGCCAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGT 180

Qy       392 CCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAA 451

Db       181 CCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGACCCTGAAG 240

Qy       452 AGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGAT 511

Db       241 AGGAAGCTTGGCGGGAGCGGTTGTGCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGAT 300

Qy       512 GCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCG 571

Db       301 GCTCACTTCTGTGCCGTCTGCAGTGATTATGCATCTGGGTATCATTACGGTGTCTGGTCC 360

Qy       572 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGT 631

Db       361 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTATATCTGT 420

Qy       632 CCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGA 691

Db       421 CCAGCCACGAATCACGTGTACGATAGACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGA 480

Qy       692 CTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGG 751

Db       481 CTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGCATCCAGAGACGCAAAGGTCTGCC 540

Qy       752 TACCGCCTTCTGCGCCAGACAGAAGAACTGCCGAGGAGCAGTGCAGTGTGCGCGCAAGGAC 811

Db       541 TACCGAATAGTACGAAGACAGAGAAGTGCCAGGAGCAGCGTGCATTGCCTGAACAAAGCC 600

Qy       812 AAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGAGGCCTGAGCCCC 871

Db       601 AAGAGAACCACTGGGCACACCCCGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCC 660

Qy       872 GAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCC 931

Db       661 GAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCC 720

Qy       932 AGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAG 991

Db       721 AGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAA 780

Qy       992 TTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTC 1051

Db       781 CTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTC 840

Qy      1052 GACCAAGTCCGCCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGAGGGGTCATGTCC 1111

Db       841 GACCAAGTCCGCCTCTTGGAAAGCTGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTCC 900

Qy      1112 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGAT 1171

Db       901 CGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGAT 960

Qy      1172 GAGGCGAAATGCCTAGAAGGAATTCTGGAAATCTTTGACATGCTGCTGGCAACTACTTCA 1231

Db       961 GAGGGGAAGTGCCTGGAAGGCATTCTGGAAATCTTTGACATGCTGCTGGCCACCACGTCA 1020

Qy      1232 AGGTTTCGAGAGTTAAAAGTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1291

Db      1021 CGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTCTGTGTGAAGGCCATCATTCTC 1080

Qy      1292 CTCAATTCCAGTATGTACCGTCTGGTCACAGCGACCCAGGATGCTGCACAGCAGCGGGAAG 1351

Db      1081 CTCAACTCCAGTATGTACCACTTGGCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAG 1140

Qy      1352 CTGGCTCACTTGCTGAACGCCGTCACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGC 1411

Db      1141 CTCACACACCTATTGAACGCAGTGACAGATGCCCTGGTCTGGGTCATTTCCAACAGTAGA 1200

Qy      1412 ATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTC 1471

Db      1201 ATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTC 1260

Qy      1472 AGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTC 1531

Db      1261 AGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTC 1320

Qy      1532 CCACTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCC 1591

Db      1321 CCGGTGTACGACCTGCTGCTGGAGATGCTGAATGCTCACACGCTTCGAGGGTACAAGTCC 1380

Qy      1592 TCCATCACGGGGTCGGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAG 1651

Db      1381 TCAATCTCGGGGTCTGGGTGCTGCTCGACAGAGGACAGTAAGAGCAAAGAGGGCTCCCAG 1440

Qy      1652 AACCCACAGTCTCAGTGA 1669

Db      1441 AACCTCCAGTCTCAGTGA 1458
```

```
RESULT 24
AAF59900
ID   AAF59900 standard; DNA; 640 BP.
XX
AC   AAF59900;
XX
DT   04-MAY-2001  (first entry)
XX
DE   EGFP/human oestrogen receptor beta (ER-beta) fusion gene fragment.
XX
KW   Human oestrogen receptor beta; EGFP/ER-beta fusion gene;
KW   enhanced green fluorescent protein; labelled hormone receptor;
KW   optical signal; ligand binding; hormonal disorder;
KW   sex hormone associated; ovulation; drug screening; drug discovery;
KW   ds.
XX
OS   Chimeric - Aequorea victoria.
OS   Chimeric - Homo sapiens.
XX
PN   WO200107919-A1.
XX
PD   01-FEB-2001.
XX
PF   24-JUL-2000; 2000WO-JP04930.
XX
PR   23-JUL-1999;  99JP-0209860.
PR   31-MAY-2000; 2000JP-0163475.
PR   31-MAY-2000; 2000JP-0163476.
XX
PA   (OLYU ) OLYMPUS OPTICAL CO LTD.
XX
PI   Sakamoto H,  Kato N;
XX
DR   WPI; 2001-168581/17.
XX
PT   Determination of the interaction of a substance for investigation and
PT   diagnosis of hormonal disorders using an optically labelled receptor
PT   protein -
XX
PS   Disclosure; Fig 9; 75pp; Japanese.
XX
CC   The invention relates to a method for examining the interaction of a
CC   test substance with a hormone receptor protein. The method involves
CC   contacting a test substance with a hormone receptor protein (e.g., an
CC   oestrogen receptor) which is labelled with an optical marker (e.g., a
CC   fluorescent protein), where the marker is capable of undergoing a change
CC   in its optical properties when the receptor binds a ligand. The optical
CC   signal generated in the presence of the test compound is then compared
CC   with that generated in the absence of the test compound to determine
CC   whether the test compound is interacting with the hormone receptor
CC   protein. The invention also encompasses the labelled hormone receptor
```

```
CC   protein, the gene encoding it, and vectors and host cells comprising
CC   such genes. The method of the invention is used for the investigation
CC   and diagnosis of hormonal disorders, especially those associated with
CC   sex hormones e.g., suppression of ovulation. The present sequence
CC   represents a fragment (containing the junction region) of an EGFP
CC   (enhanced green fluorescent protein)/human oestrogen receptor beta
CC   (ER beta) fusion gene constructed in the invention.
XX
SQ   Sequence 640 BP; 166 A; 169 C; 142 G; 163 T; 0 other;

  Query Match             30.7%;  Score 584;  DB 22;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 5e-144;
  Matches  584;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          31 TTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAAAAA 90
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          58 TTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAAAAA 117

Qy          91 CTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACCCCT 150
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         118 CTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACCCCT 177

Qy         151 GGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCCAGC 210
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         178 GGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCCAGC 237

Qy         211 CATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTT 270
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         238 CATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTT 297

Qy         271 GGAAGGTGGGCCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCA 330
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         298 GGAAGGTGGGCCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCA 357

Qy         331 CCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAG 390
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         358 CCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAG 417

Qy         391 TCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAA 450
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         418 TCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAA 477

Qy         451 AAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 510
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         478 AAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 537

Qy         511 TGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTC 570
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         538 TGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTC 597

Qy         571 GTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGA 613
               |||||||||||||||||||||||||||||||||||||||||||
Db         598 GTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGA 640


RESULT 24
ABA61882
ID   ABA61882 standard; DNA; 597 BP.
XX
AC   ABA61882;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #10187.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
```

```
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180412.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE ) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 1; SEQ ID NO 10187; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   foetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp wipo int/pub/published_pct sequences.
XX
SQ   Sequence 597 BP; 168 A; 152 C; 105 G; 172 T; 0 other;

  Query Match             24.1%;  Score 438;  DB 22;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-105;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          17 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 76

Qy          61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          77 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 136

Qy         121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         137 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 196

Qy         181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         197 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 256

Qy         241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         257 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 316

Qy         301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         317 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 376

Qy         361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         377 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 436

Qy         421 CACCTTACCTGTAAACAG 438
               ||||||||||||||||||
Db         437 CACCTTACCTGTAAACAG 454


RESULT 25
ABA29437
ID   ABA29437 standard; DNA; 597 BP.
```

```
XX
AC   ABA29437;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #7903 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236459.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 1; SEQ ID No 7903; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 597 BP; 168 A; 152 C; 105 G; 172 T; 0 other;

  Query Match             23.1%;  Score 438;  DB 22;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-105;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 76

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 136

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 196

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 256

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 316

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 376

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 436

Qy        421 CACCTTACCTGTAAACAG 438
              ||||||||||||||||||
Db        437 CACCTTACCTGTAAACAG 454


RESULT 26
AAK10196
ID   AAK10196 standard; DNA; 597 BP.
XX
AC   AAK10196;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 10187.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 10187; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 597 BP; 168 A; 152 C; 105 G; 172 T; 0 other;

  Query Match             23.1%;  Score 438;  DB 22;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-105;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          17 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 76

Qy          61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          77 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 136

Qy         121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         137 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 196

Qy         181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         197 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 256

Qy         241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         257 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 316

Qy         301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         317 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 376

Qy         361 ATCACATCTGTATGCCGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         377 ATCACATCTGTATGCCGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 436

Qy         421 CACCTTACCTGTAAACAG 438
                    ||||||||||||||||||
Db         437 CACCTTACCTGTAAACAG 454


RESULT 27
AAK46093
ID   AAK46093 standard; DNA; 597 BP.
XX
AC   AAK46093;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 10650.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0642366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE ) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 10650; 658pp - Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 597 BP; 168 A; 152 C; 105 G; 172 T; 0 other;

  Query Match             23.1%;  Score 448;  DB 22;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-105;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          17 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 76

Qy          61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          77 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 136

Qy         121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         137 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 196

Qy         181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         197 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 256

Qy         241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         257 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 316

Qy         301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         317 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 376

Qy         361 ATCACATCTGTATGCCGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         377 ATCACATCTGTATGCCGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 436

Qy         421 CACCTTACCTGTAAACAG 438
                    ||||||||||||||||||
Db         437 CACCTTACCTGTAAACAG 454


RESULT 28
AA141808
ID   AA141808 standard; DNA; 597 BP.
XX
AC   AA141808;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #10494 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0642366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
```

```
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 10494; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENs).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 597 BP; 168 A; 152 C; 105 G; 172 T; 0 other;

  Query Match             23.1%;  Score 438;  DB 22;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-105;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 76

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 136

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 196

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATCAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATCAA 256

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 316

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 376

Qy        361 ATCACATCTGTATGCGGAACCTCAAAACAGTCCCTGGTGAAGCAAGATCCCTACAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 ATCACATCTGTATGCGGAACCTCAAAACAGTCCCTGGTGAAGCAAGATCCCTACAACA 436

Qy        421 CACCTTACCTGTAAACAG 438
              ||||||||||||||||||
Db        437 CACCTTACCTGTAAACAG 454


RESULT 29
ABS10245
ID   ABS10245 standard; DNA; 597 BP.
XX
AC   ABS10245;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human genome-derived single exon probe from lung SEQ ID No 10236.
XX
KW   Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US00665.
XX
PR   04-FEB-2000; 2000US-180312P.
PR   26-MAY-2000; 2000US-207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   02-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-234687P.
PR   27-SEP-2000; 2000US-236359P.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2002-114183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples -
XX
PS   Claim 1; SEQ ID No 10236; 634pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of
CC   probes; the novel set of probes which hybridise at high stringency to a
CC   nucleic acid expressed in the human lung; measuring gene expression in a
CC   sample derived from human lung, comprising (a) contacting the array with
CC   a collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of
CC   the array; identifying exons in a eukaryotic genome, comprising
CC   (a) algorithmically predicting at least one exon from genomic sequences
CC   of the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene
CC   expression analysis, and for identifying exons in a gene, particularly
CC   using human lung derived mRNA and for the study of lung diseases
CC   such as asthma, lung cancer, chronic obstructive pulmonary disease
CC   (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC   fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC   Niemann Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC   haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC   pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC   pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC   and hyaline membrane disease. The present sequence is a single exon
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 597 BP; 168 A; 152 C; 105 G; 172 T; 0 other;
```

```
  Query Match             23.1%;  Score 438;  DB 24;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-105;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 76

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 136

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 196

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 256

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 316

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 376

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 436

Qy        421 CACCTTACCTGTAAACAG 438
              ||||||||||||||||||
Db        437 CACCTTACCTGTAAACAG 454


RESULT 30
AAS43104
ID   AAS43104 standard; DNA; 825791 BP.
XX
AC   AAS43104;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human oestrogen receptor beta gene.
XX
KW   Human; oestrogen receptor beta; ERbeta; ds; SNP; chromosome 6q.25.1;
KW   single nucleotide polymorphism; cardiovascular disease;
KW   autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;
KW   osteoarthritis; osteoporosis; breast cancer; endometrial cancer.
XX
OS   Homo sapiens.
XX
PN   WO200162793-A2.
XX
PD   30-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05360.
XX
PR   22-FEB-2000; 2000US-0183755.
PR   24-JAN-2001; 2001US-0768185.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;
XX
DR   WPI; 2001-582041/65.
DR   P-PSDB; AAU27322.
XX
PT   Oestrogen receptor gene and protein polymorphisms useful for diagnosis
PT   of individuals at risk of developing bone disorders
XX
PS   Example 2; Figure 1; 245pp; English.
XX
CC   The invention relates to a novel isolated peptide comprising or
CC   consisting of an amino acid sequence selected from an amino acid sequence
CC   of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
CC   10 amino acids), antibodies against them, nucleic acids encoding
CC   them (including vectors for transforming cells). The gene for human
CC   ERbeta is located on chromosome 6q.25.1. The variants are encoded
CC   by single nucleotide polymorphisms (SNP). The variant peptides and
CC   proteins can be used in assays to determine the biological
CC   activity of the protein, to raise antibodies, as a reagent in assays
CC   designed to quantitatively determine levels of the protein in
CC   biological fluids, to identify compounds that modulate receptor
CC   activity and to screen compounds for the ability to stimulate or
CC   inhibit interaction between the receptor protein and a target molecule
CC   that normally interacts with the receptor protein e.g. oestrogen.
CC   The antibody can be used to isolate the protein, to assess expression in
CC   disease states e.g. cardiovascular disease and autoimmune disease (e.g.
CC   systemic lupus erythematosus, arthritis, rheumatism and osteoarthritis),
CC   osteoporosis, breast cancer and endometrial cancer. In addition
CC   the antibodies can be used in pharmacogenomic analysis and inhibiting
CC   protein function, e.g. blocking the binding of the oestrogen receptor
CC   protein to a binding partner such as a ligand. The nucleic acids
CC   encoding the proteins can be used as probes, primers, chemical
CC   intermediates and in biological assays. The present sequence is the
CC   human ERbeta gene.
XX
SQ   Sequence 825791 BP; 94098 A; 68292 C; 67970 G; 95431 T; 0 other;

  Query Match             23.1%;  Score 438;  DB 22;  Length 825791;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-104;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104460 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 104519

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104520 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 104579

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104580 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 104639

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104640 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 104699

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104700 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 104759

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104760 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 104819

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104820 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 104879

Qy        421 CACCTTACCTGTAAACAG 438
              ||||||||||||||||||
Db     104880 CACCTTACCTGTAAACAG 104897


RESULT 31
AAD41046
ID   AAD41046 standard; cDNA; 2764 BP.
XX
AC   AAD41046;
XX
```

```
DT   18-JUN-2002  (first entry)
XX
DE   Medaka fish oestrogen receptor cDNA.
XX
KW   Medaka fish; oestrogen; receptor; transgenic; thrombosis; therapeutic;
KW   endocrine disrupting chemical; bioassay system; ss
XX
OS   Oryzias latipes.
XX
FH   Key             Location/Qualifiers
FT   CDS             211..1938
FT                   /*tag= a
FT                   /product= "Medaka fish oestrogen receptor protein"
XX
PN   EP1180684-A1.
XX
PD   20-FEB-2002.
XX
PF   27-JUN-2001; 2001EP-0115472.
XX
PR   17-AUG-2000; 2000JP-0247729.
XX
PA   (UYHI-) UNIV HIROSHIMA.
XX
PI   Yamashita I;
XX
DR   WPI; 2002-229967/29.
DR   P-PSDB; AAE19430.
XX
PT   New polynucleotide from Oryzias latipes, useful for producing
PT   transgenic medaka fish having sensitivity to a very small amount of
PT   estrogen, which is useful for testing estrogen-like endocrine
PT   disrupting chemicals -
XX
PS   Claim 1; Page 9-16; 23pp; English.
XX
CC   The present sequence is medaka fish (Oryzias latipes) oestrogen receptor
CC   cDNA. The polynucleotide is useful for producing transgenic medaka fish
CC   having sensitivity to a very small amount of oestrogen or a medaka fish
CC   having one or more thrombi. The highly oestrogen-sensitive medaka fish
CC   is useful for testing oestrogen-like endocrine disrupting chemicals,
CC   for elucidating development mechanism of thrombosis, or as a bioassay
CC   system for developing a therapeutic agent for thrombosis.
XX
SQ   Sequence 2764 BP; 649 A; 736 C; 725 G; 654 T; 0 other;

  Query Match             20.9%;  Score 395.8;  DB 24;  Length 2764;
  Best Local Similarity   62.9%;  Pred. No. 3.3e-94;
  Matches  704;  Conservative    0;  Mismatches  367;  Indels   48;  Gaps    4;

Qy        509 GATGCTCACTTCTGCGGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGG 568
              ||   | |  ||||||| || |||||||| ||||| || || || || |||||||| || |||
Db        619 GACACGCGTTTCTGCGGCCGTCTGCAGCGACTACGCCTCTGGGTACCACTATGGGGTCTCC 678

Qy        569 TCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGCACATAATGATTATATT 628
              || ||||| || || |||||||| || || || ||||| || || || ||||| |||||
Db        679 TCTTGTGAGGGCTGCAAGGCCTTCTTCAAGAGGAGCATCCAGGGTCACAATGACTATATG 738

Qy        629 TGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCTGC 688
              || ||||| || |||||||| || || || | ||| ||| | ||||||||||||||| ||
Db        739 TGCCCAGCGACCAATCAGTGCACTATTGACACAAATCGCAGGAAGAGCTGCCAGGCTTGT 798

Qy        689 CGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTCTCC---------------------- 729
              || ||| ||||||||||||||||||||||||| ||||  | ||
Db        799 CGTCTTAGGAAGTGTTACGAAGTGGCAATGATGAAAGGCGGTGTGCGCAAGGACCGCATT 858

Qy        730 -----CTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTCCGGAGACAGAGAAGTGCCAC 784
                  | ||| | || | | |  ||    || |||      || | |    ||   |
Db        859 CGCATTTTACGGCGTGACAAACGGCGGACAGGCGTTGGTGATGGAGACAAGGTTGTAAAG 918

Qy        785 GAGCAGCTGCACTGTGCCG---------------GCAAGGCCAAGAGAAGTGGCGGCCAC 829
              |  |||  |||    | |               | ||   ||  || |  || ||
Db        919 GGTCAGGAGCATAAAACGGTGCATTATGATGCAAGCAAACGCAGCAGCAGCAGGAGGAGA 978

Qy        830 GCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACC 889
              |     |||  | |     ||| |  ||   |  ||     ||| ||||| |
Db        979 GGAGGAGGAGGAGGAGGAAGACTGTCTGTGACCAGCATACCTCCTGAGCAGGTGCTGCTC 1038

Qy        890 CTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGC------CCAGTGCGCCCTTC 943
              ||||| ||| |    ||||    | ||   |   || |         ||    || | |
Db        1039 CTCCTTCAGGGCGCCGAGCCCCCGATACTCTGCTCGGGTCAGAAGTTCAGCCGACGTAC 1098

Qy        944 ACCGAGGCCTCCATCATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATG 1003
              |||||| | |||||||||     || ||||   |||| |||||||| |||| ||||||
Db        1099 ACCGAGGTCACCATGATGACCCTGCTCACCAGCATGGCAGACAAGGAGCTGGTCCACATG 1158

Qy        1004 ATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGG 1063
               |||  |||||||||||||| | || || ||| || ||||   ||||  ||| || |||| |
Db        1159 ATCGCCTGGGCCAAGAAGCTCCCAGGTTTTGTGCAGCTGTCCCTGCACGATCAGGTGCTG 1218

Qy        1064 CTCTTGGAGAGCTCTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGAC 1123
               ||  |||||||||  ||| |||||||| | ||||| || || || ||| | || ||  ||
Db        1219 CTGCTGGCAGAGCTCGTGGCTGGAGGTGCTCATGATCGGCCTCATTTGGAGGTCCATCCAC 1278

Qy        1124 CACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGC 1183
                ||||| ||||||||||||||| | ||| ||  | |||||||| |||||||  | |||
Db        1279 TGTCCCGGGAAGCTCATCTTTGCACAAGACCTCATCCTGGACAGGAATGAGGGAGACTGC 1338

Qy        1184 GTAGAAGGAATTCTGGAAATCTTTGACATGCTGCTGGCAACTACTTCAAGGTTTCGAGAG 1243
               || ||||| ||   ||| ||||| |||||||| ||||| ||| |||| | || || | |
Db        1339 GTGGAAGGCATGACGGAGATCTTCGACATGCTGCTGGCCACTGCTTCCCGCTTCCGTGTG 1398

Qy        1244 TTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATCATCCTGCTCAATTCCAGT 1303
                | |||||| |||   | ||||   |||| |||| || || ||  | ||||| ||| ||
Db        1399 CTCAAACTCAAACCTGAGGAATCGTCTGCCTCAAAGCTATTATTTTACTCAACTCCGGT 1458

Qy        1304 ATGTACCCTCTGGTCACAGCGACCCAGGATGC---TGACAGCAGCCGGAAGCTGGCTCAC 1360
                |   || |   ||| |  ||| ||| |   | ||| |||| |  | |      |
Db        1459 GCTTTTTCTTTCTGCAGCGGCACCATGGAGCCACTTCACAACAGCGCGGCGGTTCAGAGC 1518

Qy        1361 TTCCTGAACGCCCTCACCGATGCTTTGGTTTGGGTGATTGCCAAGACCGGCATCGCTTCC 1420
               ||||| || || | || || ||  |  ||       ||     ||   ||   ||  ||
Db        1519 ATGCTGGACACCATCACAGACGCACTCATTCATTACATCAGTCAGTCGGGTTACTTGGCC 1578

Qy        1421 CAGCAGCAATCCATGCGGCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGGC 1480
               ||| |||  | |  || | ||| | || ||  |||| || |||||| |||||||   |
Db        1579 CAGGAGCAGGCGACACGGCAGCCCCAGCTGCTCCTGCTGCTCTCCCACATCAGGCACATG 1638

Qy        1481 AGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTAT 1540
               || ||||| |||||||| || ||   || ||||||||||| ||    |||||  | |||
Db        1639 AGCAACAAAGGCATGGAGCACCTCTACAGCATGAAGTGCAAGAACAAAGTCCCTCTTTAT 1698

Qy        1541 GACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGC 1579
               ||||| || |||||||||||  |||||||   || | |
Db        1699 GACCTCCTACTGGAGATGCTCGATGCCCACCGCCTGCAC 1737


RESULT 32
AAF29925
ID   AAF29925 standard; DNA; 1374 BP.
XX
AC   AAF29925;
XX
DT   04-APR-2001  (first entry)
XX
DE   Human estrogen receptor alpha isoform #2 DNA.
XX
KW   Human; estrogen receptor alpha; cancer; osteoporosis; bone;
KW   Alzheimer's; cardiovascular; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100823-A1.
XX
PD   04-JAN-2001.
XX
```

```
PF   27 JUN 2000; 2000WO-EP005981.
XX
PR   29 JUN 1999;   99IT-MI01434.
XX
PA   (EUMO ) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Gannon F,  Denger S,  Flouriot G;
XX
DR   WPI; 2001-137959/14.
XX
PT   Novel isoforms of human estrogen receptor alpha useful for preparing
PT   therapeutic agents for treating cancer, osteoporosis, Alzheimer's
PT   disease and cardiovascular diseases -
XX
PS   Claim 1; Page 44-45; 53pp; English.
XX
CC   The present invention relates to a human estrogen receptor
CC   (hER)-alpha isoform. Molecules which modulate the activity of
CC   the estrogen receptor are useful for the preparation of therapeutic
CC   agents for treating cancer, osteoporosis and other bone disorders,
CC   Alzheimer's disease and cardiovascular diseases.
XX
SQ   Sequence 1374 BP; 349 A; 346 C; 386 G; 293 T; 0 other;

  Query Match             20.6%;  Score 390.8;  DB 22;  Length 1374;
  Best Local Similarity   62.4%;  Pred. No. 5e-93;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | ||   ||| || ||| || || ||||  || || || || || || || |||||||||
Db        124 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 183

Qy        566 TGGTCGTGTGGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| |||| || || |||||||| || || |||| ||||||||||||| || |||
Db        184 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 243

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db        244 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 303

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||| |||| |||| || | ||| ||
Db        304 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 363

Qy        746 TGTGG--------  --GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
                | ||          | | | |     | |||| |    | |  | | |
Db        364 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 423

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCCGCCCCGAGTGCG-------- 843
                | | |    | || ||||  || |  ||   ||| |   ||
Db        424 GTGGGTCTGCTGGAGACATGAGAGCTGCCAACCCTTGGCCAAGCCCGCTCATGATCAAA 483

Qy        844 -    ----  --GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                 | |    |||       ||||||   |||| ||| | | ||     || |
Db        484 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 543

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCCGCCCTTCACCGAG 949
              |||| |||||||| |||    |       |          || |     ||||||| ||
Db        544 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 603

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAAC 1009
              || || ||||||   |  |||||||| |||| |||| |||| |||| ||||||||| |
Db        604 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 663

Qy       1010 TGGGCCAAGAGATTCCCGGCTTTGTGGAGCTCACCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || |||||||||||  | | |||    || || || |  ||  |
Db        664 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 723

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||   | |||||||| || || |||||
Db        724 GAATGTGCCTGGCTAGACATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 783

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAA 1189
```

```
              || ||||  | ||||||||  |  |  |  |||||||| |  |||| |||| ||||
Db        784 GGGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 843

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| |||| |||||||| |||| || |  | ||  |||| ||   | | ||
Db        844 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 903

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | || |||| | || ||  |||| |||||  |  |||||
Db        904 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 963

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  | |  |    |||| | |  ||   ||    |    |  |    |  | |||
Db        964 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1023

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |     |||| ||||||   ||| | ||   ||||||
Db       1024 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1083

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATCGGAGTAAC 1486
              ||   |  ||| ||||  | ||||  |  |||| |||||| |||||||   |||||||
Db       1084 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1143

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || |||||||||||| || |||| ||  | |||||||||
Db       1144 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1203

Qy       1547 CTGCTGGAGATGCTGGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||  | ||||||   || |  | |  ||
Db       1204 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 1243


RESULT 33
ABL57497
ID   ABL57497 standard; cDNA; 1788 BP.
XX
AC   ABL57497;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human nuclear oestrogen receptor alpha coding sequence.
XX
KW   oestrogen; receptor; human; transgenic mouse; cytostatic;
KW   antiinflammatory; antidiabetic; endocrine; anorectic; hepatotropic;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1788
FT                   /*tag= a
FT                   /product= "Nuclear oestrogen receptor"
XX
PN   WO0200228175-A2.
XX
PD   11-APR-2002.
XX
PF   28-SEP-2001; 2001WO-IB02246.
XX
PR   03-OCT-2000; 2000FR-0012570.
PR   11-MAY-2001; 2001US-0853033.
XX
PA   (ADER-) ADEREGEM ASSOC DEV RECH EN GENETIQ.
XX
PI   Chambon P,  Metzger D;
XX
DR   WPI; 2002-464217/49.
DR   P-PSDB; ABB76378.
XX
PT   A transgenic mouse, useful in screening for medicaments for the
PT   treatment of e.g., diabetes or skin cancers, comprises a fusion protein
PT   between a recombinase Cre, and a modified ligand binding domain of the
PT   nuclear estrogen receptor alpha -
```

```
XX
PS   Disclosure; Page 128-130; 149pp; English.
XX
CC   The present sequence is the coding sequence of the human nuclear
CC   oestrogen receptor alpha (I).  The invention relates to a non-human
CC   metazoan organism, especially a transgenic mouse, characterised in
CC   that at least one cell comprises: (i) a fusion protein formed from
CC   a recombinase (Cre), a hinge region (preferably derived from human
CC   (I)), and a modified ligand binding domain (LBD) of a nuclear
CC   oestrogen receptor, such as (I), its fragment or variant; and (ii)
CC   one or more genes or DNA sequences of interest belonging to the
CC   genome of the organism, into which one or more recognition sites
CC   of the recombinase protein are inserted.  The (I) LBD domain is
CC   preferably modified by a G521R, G400V, or M543A/L544A mutation.
CC   The fusion protein has negligible, or even zero, recombinase
CC   activity in the presence of a natural ligand such as oestradiol,
CC   but recombinase activity is induced by a small quantity of a
CC   synthetic ligand that has antioestrogenic activity, e.g. tamoxifen
CC   or 4-hydroxytamoxifen.  The metazoan organism or its cells, such
CC   as epidermal cells, hepatocytes or adipocytes, are useful in
CC   carrying out a spatiotemporally controlled site-specific
CC   recombination of a DNA sequence of interest in its natural
CC   chromatin environment.  It is also used in screening of medicaments
CC   for pathological conditions associated with an alteration of the
CC   expression and/or function of the DNA sequence of interest, such as
CC   skin cancer, inflammation, diabetes, alopecia, obesity, or in
CC   promoting hepatic regeneration.
XX
SQ   Sequence 1788 BP; 418 A; 515 C; 508 G; 347 T; 0 other;

  Query Match             20.6%;  Score 390.8;  DB 24;  Length 1788;
  Best Local Similarity   62.4%;  Pred. No. 5.7e-93;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        538 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 597

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| |||| || || |||||||| || || |||| |||||||||||||| || |||
Db        598 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 657

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db        658 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 717

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGCAATGGTGAAGTGTCGCTCCCGGACAGACAGA 745
              ||||| || || || || |||||||||||||| |||| |||| || | |||  ||
Db        718 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 777

Qy        746 TGTGG----------GTACCGCTTCTCTCCGGAGACACAGAAGTGCCGACGAGCAGCTG 795
              | ||          | | | |    | |||| |     | | | | |  |
Db        778 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGTGAA 837

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 843
                | | |     | || |||| || | ||    ||| |   ||
Db        838 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 897

Qy        844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTACTCCTCACCCTC 892
                       | |    |||       |||||   |||| ||| | ||     || |
Db        898 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 957

Qy        893 CTGGAGGCTGAGCCGCCCCATGIGCTGATC---AGCCGCCCCAGTGCCCCTTCACCCAG 949
              |||| |||||||| |||    |     |          || |     |||||||  ||
Db        958 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATGCCTACCAGACCCTTCAGTCAA 1017

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  |||||||| |||| |||| |||| |||| ||||||||| |
Db       1018 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1077

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || |||||||||||  | | |||    || || || |  ||  |
Db       1078 TGGGCCAAGAGGGTGCCAGGCTTTGTGCATTTGACCCTCCATGATCAGCTCCACCTTCTA 1137

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              || |   ||| | ||| |  | ||||| || ||  | ||||||| || || |||||
Db       1138 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1197

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAA 1189
              || ||||  | ||||||||  |  |  |  |||||||| |  |||| |||| ||||
Db       1198 GGGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1257

Qy       1190 CGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGHTTCGAGAGTTAAAA 1249
              || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db       1258 GGCATGGTGGACATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1317

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||  | || ||  |||| |||||  |  |||||
Db       1318 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1377

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              | | | | |    |||| | |  ||  ||   |    | | |   | | |||
Db       1378 ACATTTCTGTCCAGCAGCCTGAAGTCTCTGGAAGAGAAGGACCATATCTCACCGAGTCCTG 1437

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||||  |    |||| ||||||   ||| |  ||   ||||||
Db       1438 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1497

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||    | ||| |||||  | |||||  |  |||| ||||||  |||||||   |||||||
Db       1498 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1557

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || ||||||||||||| || ||||| ||  | |||||||||
Db       1558 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1617

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||| | ||||||   || |  | |  ||
Db       1618 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 1657


RESULT 34
AAT06873
ID   AAT06873 standard; DNA; 4963 BP.
XX
AC   AAT06873;
XX
DT   13-MAR-1996  (first entry)
XX
DE   PhCMV*-1 promoter and human oestrogen receptor.
XX
KW   Transactivator; tTA; tet operator; tetO; gene expression;
KW   tetracycline-responsive promoter; PhCMV*-1; human cytomegalovirus;
KW   transcription activation; eukaryotic cell; oestrogen receptor;
KW   ds; ss.
XX
OS   Chimeric synthetic;
OS   Chimeric Homo sapiens.
XX
PN   US5464758-A.
XX
PD   07-NOV-1995.
XX
PF   14-JUN-1993;   93US-0076726.
XX
PR   14-JUN-1993;   93US-0076726.
XX
PA   (BUJA/) BUJARD H.
PA   (GOSS/) GOSSEN M.
XX
PI   Bujard H, Gossen M;
XX
DR   WPI; 1995-392612/50.
XX
PT   Polynucleotide encoding trans:activator fusion protein contg. tet
PT   repressor    used to control expression of gene regulated by minimal
```

```
PT   promoter linked to tet operon, and vectors and cells where gene
PT   expression is regulated by tetracycline
XX
PS   Disclosure; Fig 10A-B; 37pp; English.
XX
CC   cDNA coding for the human oestrogen receptor was placed under
CC   control of the PhCMV*-1 promoter (see AAT06869), resulting in
CC   tetracycline regulated expression vector pUHDppr-3 contg. the
CC   sequence given in AAT06873. The construct was transfected into
CC   eukaryotic cells expressing the tTA transactivator (AAT06867) to
CC   allow tetracycline-regulated expression of the receptor.
XX
SQ   Sequence 4964 BP; 1259 A; 1300 C; 1278 G; 1126 T; 0 other;

  Query Match             20.6%;  Score 390.8;  DB 16;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 9e-93;
  Matches  699;  Conservative    0;  Mismatches  482;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db       1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy        566 TGGTCGTGTGAAGGATGTAAGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || ||||||||| || | ||||||||||||||||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCCCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | ||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        746 TGTGG  --------- GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||             | | | |      | |||| |    | |  | | |
Db       1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG--    - --- 843
              | | |     | ||| ||||  || |  ||    ||| |   ||
Db       1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        844            -GGAGCTGCTGCTGGACGCCCTGACGCCCCGAGCAGCTAGTGCTCACCCTC 892
                          | |    |||       ||||||   |||| ||| | ||  |  || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1422

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---ACCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||    |       |          || |     |||||||  ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  |||||||| |||| |||| |||| |||| ||||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              || |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAACGTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  ||||  | |||||||| |  | |  |  |||||||| |  |||| ||||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||| | || ||  |||| |||| |  ||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  | | |  |    |||| | |  ||   ||     |    | |   |  | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTCACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
                ||    | || ||  |||||  |     |||| ||||||   ||| |  ||   ||||||
Db       1903 CACAACATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || ||||||||||||| || ||||| ||  | |||||||||
Db       2023 AAAGCCATGCAGCATCTGTACAGCATGAAGTGCAACAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 45
AAQ76270
ID   AAQ76270 standard; DNA; 4964 BP.
XX
AC   AAQ76270;
XX
DT   17-JUL 1995  (first entry)
XX
DE   PhCMV*-1 controlled progesterone receptor.
XX
KW   tTA; transactivator; tetracycline-controllable transactivator;
KW   conditional inactivation; homologous recombination; gene expression;
KW   gene regulation; gene therapy; tetracycline-resistance; tetR;
KW   transgenic animal; PhCMV*-1; promoter; teto; CMV;
KW   progesterone receptor; ds.
XX
OS   Human cytomegalovirus Towne.
XX
PN   WO9429442-A.
XX
PD   22-DEC-1994.
XX
PF   14-JUN-1994;   94WO-US06734.
XX
PR   14-JUN-1993;   93US-0076327.
XX
PA   (BADI ) BASF AG.
XX
PI   Bujard H,  Gossen M,  Salfeld JG,  Voss JW;
XX
DR   WPI; 1995-036472/05.
XX
PT   Regulatory systems using tetracycline-controllable transactivator
PT   (tTA) - useful for conditional inactivation or modulation of
PT   gene expression in a host cell or animal
XX
PS   Disclosure; Page 54-58; 104pp; English.
XX
CC   The minimal promoter PhCMV*-1 is based on the human CMV
CC   intermediate-early promoter (with enhancer region removed), fused
CC   to a tetracycline operator, teto. The promoter is used to
CC   express tTA transactivator in host cells. In clone pUHD BGR4
CC   (given in AAQ76269), cDNA for rabbit progesterone receptor is under
CC   the control of PhCMV*-1.
XX
SQ   Sequence 4964 BP; 1259 A; 1298 C; 1280 G; 1126 T; 0 other;

  Query Match             20.6%;  Score 390.8;  DB 16;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 9e-93;
  Matches  699;  Conservative    0;  Mismatches  482;  Indels   39;  Gaps    4;
```



RESULT 36
AAX81725
ID AAX81725 standard; DNA; 4963 BP.
XX
AC AAX81725;
XX
DT 27-AUG-1999 (first entry)
XX
DE cDNA encoding rabbit progesterone receptor under CMV promoter.
XX
KW Tetracycline controllable transactivator; tTA; transgenic mouse;
KW transgene; gene function; gene expression; knockout animal;
KW conditional inactivation; promoter; progesterone receptor; ds.
XX
OS Synthetic.
OS Human cytomegalovirus.
OS Oryctolagus cuniculus.
XX
PN US5922927-A.
XX
PD 13-JUL-1999.
XX
PF 21-JUL-1997; 97US-0897719.
XX
PR 14-JUN-1994; 94US-0260452.
PR 14-JUN-1993; 93US-0076327.
PR 21-JUL-1997; 97US-0897719.
XX
PA (BADI) BASF AG.
XX
PI Bujard H, Gossen M, Salfeld JG, Voss JW;
XX
DR WPI; 1999-404496/34.
XX
PT Production of tetracycline regulated transgenic mice, useful for
PT studying gene expression
XX
PS Disclosure; Fig 10A-B; 66pp; English.
XX
CC The present sequence represents a cDNA encoding rabbit progesterone
CC receptor under the control of a human cytomegalovirus promoter.
CC It is used to create the transgenic animals of the invention. The
CC specification describes a method for producing a transgenic mouse
CC having a transgene encoding a tetracycline-controllable transactivator
CC (tTA) integrated at a predetermined location within chromosomal
CC DNA. The transgenic mice are useful for studying gene function and
CC expression. The method allows the development of knockout animals for
CC many genes of interest and allows for conditional inactivation or
CC modulation of expression of a gene in a mouse.
XX
SQ Sequence 4963 BP; 1259 A; 1300 C; 1278 G; 1126 T; 0 other;

  Query Match 20.6%; Score 390.8; DB 20; Length 4963;
  Best Local Similarity 62.4%; Pred. No. 9e-93;
  Matches 699; Conservative 0; Mismatches 382; Indels 39; Gaps 4;

```
                 ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db        1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy         746 TGTGG- - ----- GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 798
                 | ||      | | | |   | ||| |    | |  | | |
Db        1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy         799 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG-------- -- 843
                   | | |     | || |||| || |  ||    ||| |  ||
Db        1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy         844 --------- ---GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                       | |   |||       ||||||  |||| ||| | ||  |  || |
Db        1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTCACGGCCGACCAGATGGTCATGGCCTTG 1422

Qy         893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC- -AGCCGCCCCAGTGCGCCCTTCACCGAG 949
                 |||| |||||||| |||   |     |      || |    |||||||  ||
Db        1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy         950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
                || || ||||||   |  ||||||||  |||| |||| |||| |||| ||||||||||
Db        1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy        1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
                ||||| |||| | | || ||||||||||  | | |||    || || || | |  |
Db        1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy        1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
                ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db        1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy        1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
                |  |||||  | |||||||||  |  |  |  ||||||| |  |||| ||||| |||||
Db        1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy        1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
                || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db        1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy        1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
                || ||     | || | | | ||  |||| | || ||  |||| |||||  |  |||||
Db        1783 CTGCAGGGAGACGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy        1310 CCTCTGGTCACAGCGACCCAGGATGCT-- GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
                |  |  | |   |||| | |  ||    |      |  |     |  | |||
Db        1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGACAAGGACCATATCCACCGAGTCCTG 1902

Qy        1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
                |.    | || ||  ||||| |     |||| ||||||  ||| | ||   ||||||
Db        1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy        1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
                ||   |  ||| ||||| | ||||| |  |||| |||||| |||||||   |||||||
Db        1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy        1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
                || |||||||| ||||||  || |||||||||||| || ||||| ||  | |||||||||
Db        2023 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy        1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
                |||||||||||||||| | ||||||   || |  | |  ||
Db        2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 37
AAX27906
ID   AAX27906 standard; DNA; 4963 BP.
XX
AC   AAX27906;
XX
DT   03-JUN-1999  (first entry)
XX
DE   Rabbit progesterone receptor under control of PhCMV*-1 promoter.
XX
KW   tTA transactivator; tetracycline-controllable transactivator;
KW   expression regulation; tet operator-linked gene; gene therapy;
KW   cancer treatment; viral disease; PhCMV*-1 promoter;
KW   progesterone receptor; ds.
XX
OS   Oryctolagus sp.
OS   Human cytomegalovirus.
XX
PN   US5888981-A.
XX
PD   30-MAR-1999.
XX
PF   07-JUN-1995;   95US-0479306.
XX
PR   07-JUN-1995;   95US-0479306.
PR   14-JUN-1993;   93US-0076327.
PR   14-JUN-1994;   94US-0260452.
XX
PA   (BADI ) BASF AG.
PA   (KNOL ) KNOLL AG.
XX
PI   Bujard H, Gossen M, Saffeld JG, Voss JW;
XX
DR   WPI; 1999-243263/20.
XX
PT   New method for regulating expression of a tet operator-linked gene
PT   - useful in gene therapy applications for treatment of cancer
PT   through overexpression of the tumour necrosis factor alpha gene
XX
PS   Disclosure; Fig 10; 69pp; English.
XX
CC   This sequence represents the rabbit progesterone receptor under
CC   the control of the PhCMV*-1 promoter sequence. The invention
CC   relates to a method for regulating expression of a tet operator-linked
CC   gene in a cell of a subject, using a tetracycline-controllable
CC   transactivator (tTA) transactivator. The method can be used to regulate
CC   the level of gene expression and therefore has applications in gene
CC   therapy e.g. in cancer treatment (overexpression of a cytokine gene,
CC   TNF alpha) or treatment of viral diseases by expression of trans-dominant
CC   negative viral transactivation proteins such as trans-dominant tat and
CC   rev mutants for HIV. The system allows modulation of the level of gene
CC   expression in a subject, which has not been possible in prior art
CC   systems.
XX
SQ   Sequence 4963 BP; 1259 A; 1400 C; 1278 G; 1126 T; 0 other;

  Query Match             20.6%;  Score 490.8;  DB 20;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 9e-93;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   49;  Gaps    4;

Qy         506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
                 | |||  ||| || ||| || || |||  || || || || || || || |||||||||
Db        1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy         566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
                ||||| ||||| || || |||||||| || || |||| |||||||||||||| || |||
Db        1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy         626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
                || |||||||| || || ||||| || || |||||||||  || | ||||||||||||||
Db        1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy         686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
                ||||| || || || || ||||||||||||||| |||| ||||    || | |||  ||
Db        1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy         746 TGTGG- - ----- --GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 798
                 | ||          | | | |     | |||| |    | |  | | |
Db        1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy         799 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG -- -      843
                   | | |     | || |||| || |  ||    ||| |  ||
```

```
Db       1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        844 ----------GGAGCTGCTGCTGCACGCCCTGAGCCCCGAGCAGCTACTGCTCACCCTC 892
                      | |     |||        ||||||   |||| ||| | ||   |   || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCGGACCAGATGGTCAGCGCCTTG 1422

Qy        893 CTGGAGGCTGACCCGCCCCATGTGCTGATC---AGCGCCCCAGTGCCGCCTTCACCGAG 949
              |||| ||||||| |||    |      |          || |     |||||||  ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGACTCCCTGACCAACCTGGCAGACAGGGAGCTGGTCCACATGATCAACT 1009
              || || ||||||   | |||||||  |||| |||| |||| |||| ||||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAGGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGGGCATTG 1069
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy       1070 GAGAGCTGTTGGATGGAGGTGTAATGATGGGGCTGATGTGGCGCATCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | |||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTGGTCTGGGGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCCTACAA 1189
              |  |||||  | ||||||||  |  |  |  |||||||| |  |||| |||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTTCGAGAGTTAAAA 1249
              || ||  |||| ||||| |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       1723 GGCATGCTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy       1250 CTCCAACACAAAGAATATCTCTCTTTCAAGGGCATCATCCTCCTCAATTCCACTACTGAG 1309
              || ||     | || | | ||  |||| | || ||  |||| |||||  |  ||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGTCTCAAATCTATTATTTTTCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  |  |    |||| | |  ||   ||    |      |  |    |  | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTGGGGTGATTGCCAAGAGCGGCATCTCCTCCACAGAC 1426
              ||   | ||  ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGCCTGGCTAAGCTCCTGATGCTCCTGTCCCACATCAGGCACATGGGACTAAC 1486
              ||   | ||| ||||  | ||||  |  |||| |||||| ||||||   |||||||
Db       1963 CACCACCAGGGCTTGGCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCACTGTATGACCTC 1546
              || |||||||| ||||||  || |||||||||||| || |||| ||  | |||||||||
Db       2023 AAACCCATCGAGCCATCTGTACAGCTCAACTGCAAGAACCTGCTGTTCGCCCTCTATGCCCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGGGGGTGCA 1586
              ||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGACATGCTCCACCCCACGGGCTACATCGGCAAA 2122


RESULT 38
AAX01370
ID   AAX01370 standard; DNA; 4963 BP.
XX
AC   AAX01370;
XX
DT   19-APR-1999  (first entry)
XX
DE   PhCMV*-1 rabbit progesterone fusion sequence.
XX
KW   tTA transactivator; transgenic mouse; tet operator-linked gene;
KW   tetracycline-controllable transactivator fusion protein;
KW   transcriptional regulatory element; ds.
XX
OS   Synthetic.
XX
PN   US5859310-A.
XX
PD   12-JAN-1999.
XX
PF   07-JUN-1995;   95US-0481970.
XX
PR   07-JUN-1995;   95US-0481970.
PR   14-JUN-1993;   93US-0076327.
PR   14-JUN-1994;   94US-0260452.
XX
PA   (BADI ) BASF AG.
XX
PI   Bujard H, Gossen M, Saltfeld JG, Voss JW;
XX
DR   WPI; 1999-152108/13.
XX
PT   Transgenic mouse - with tetracycline-dependent gene suppression
XX
PS   Disclosure; Fig 10; 78pp; English.
XX
CC   This sequence represents a fusion gene between the the PhCMV*-1 DNA
CC   sequence (containing the tet operator sequence and the hCMV minimal
CC   promoter) and the rabbit progesterone receptor coding sequence. The
CC   invention relates to a transgenic mouse having a transgene integrated
CC   into the genome of the mouse and also having a tet operator-linked gene
CC   in the genome of the mouse, where: the transgene comprises a
CC   transcriptional regulatory element functional in cells of the mouse
CC   operatively linked to a polynucleotide sequence encoding a
CC   tetracycline-controllable transactivator fusion protein (tTA), the fusion
CC   protein comprises a Tet repressor operably linked to a polypeptide which
CC   directly or indirectly activates transcription of the tet operator-linked
CC   gene in eukaryotic cells, the tet operator linked gene confers a
CC   detectable and functional phenotype on the mouse when expressed in cells
CC   of the mouse, the transgene is expressed in cells of the mouse at a level
CC   sufficient to produce amounts of the fusion protein that are sufficient
CC   to activate transcription of the tet operator linked gene; and in the
CC   absence of tetracycline or a tetracycline analogue in the mouse, the
CC   fusion protein binds to the tet operator-linked gene and activates
CC   transcription of the tet operator linked gene such that the tet
CC   operator-linked gene is expressed at a level sufficient to confer the
CC   detectable and functional phenotype on the mouse, where the level of
CC   expression of tet operator linked gene can be down-modulated by
CC   administering tetracycline or a tetracycline analogue to the mouse. The
CC   mice are useful as conditional knockout animals for studying human
CC   diseases.
XX
SQ   Sequence 4963 BP; 1259 A; 1300 C; 1278 G; 1126 T; 0 other;

  Query Match             20.6%;  Score 390.8;  DB 20;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 9e-93;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   49;  Gaps    4;

Qy        566 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 625
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db       1063 AAGGAGACTCGCTACTGCGCCACTGTGCAATCACTATGCTTCAGGCTACCATTATGGAGTC 1122

Qy        626 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 685
              ||||| ||||| || || |||||||| || || ||||| |||||||||||||| || |||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACACGAGGAAGAGCTGCCAGGCC 1182

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        746 TGTGG------------GTACCGCCTTGTGCCGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||            | | | |     | |||| |     | |  | |  |
Db       1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 843
                | | |     | || ||||  || |  ||    |||  |   ||
Db       1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362
```

```
Qy        844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                        | |    |||       ||||||   |||| ||| | ||  |  || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGCGCTTG 1422

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||    |       |       || |       |||||||  ||
Db       1423 TTGCATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||    |  ||||||||  |||| |||| |||| |||| ||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| ||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGCGGCCTGATGTGCCGCTCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGACATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCACATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  |||||  | ||||||||  |  |  |  ||||||||  |  |||| ||||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| ||||| ||||||||  |||||  ||  |  ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  ||||  | || ||  |||| ||||  |  |||||
Db       1783 CTCCAGGCAGACGCAGTTTGTCTCCCTCAAATCTATTATTTTCCTTAATTCCTCGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  |  |    |||| | |  ||   ||    |       |  |    |  | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |      |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGGCCTGGCTAACCTCCTGATGCTCCTGTCCCACAGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| ||||| | |||||  |  |||| |||||| |||||||   |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || ||||||||||||| || ||||| ||  | ||||||||
Db       2023 AAAGGCATGGAGCATCTGTACAGCATCAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 49
AAD09838
ID   AAD09838 standard; cDNA; 4963 BP.
XX
AC   AAD09838;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Rabbit progesterone receptor cDNA under control of PhCMV* 2.
XX
KW   Tetracycline; Tc; transactivator; tTA; herpes simplex virus; HSV; AIDS;
KW   transgenic plant; virion protein 16; VP16; gene therapy; cancer; rabbit;
KW   rheumatoid arthritis; hypopituitarism; tissue regeneration; progesterone;
KW   PhCMV* 2; ss.
XX
OS   Oryctolagus cuniculus.
XX
PN   US6252136-B1.
XX
PD   26-JUN-2001.
XX
PF   29-SEP-1998;   98US-0163269.
XX
PR   07-JUN-1995;   95US-0481970.
PR   14-JUN-1993;   93US-0076327.
PR   14-JUN-1994;   94US-0260452.
XX
PA   (BADI ) BASF AG.
PA   (KNOL ) KNOLL AG.
XX
PI   Bujard H; Gossen M; Saffeld JG; Voss JW;
XX
DR   WPI; 2001-431976/46.
XX
PT   New transgenic plants having tetracycline-regulated transcriptional
PT   regulatory systems, useful for characterizing the activity of
PT   substances or studying the absence or presence of tetracycline or
PT   tetracycline analogs.
XX
PS   Disclosure; Fig 10; 70pp; English.
XX
CC   The invention relates to a transgenic plant having a transgene
CC   integrated into the genome of the plant and also a tet operator-
CC   linked gene in the genome of the plant. The transgene comprises a
CC   transcriptional regulatory element functional in cells of the plant
CC   operatively linked to a polynucleotide sequence encoding a fusion
CC   protein, which activates transcription of the tet operator-linked gene.
CC   The fusion protein comprises a first polypeptide, which is a Tet
CC   repressor operatively linked to a second polypeptide (e.g. herpes simplex
CC   virus virion protein 16 (HSV VP16), which directly or indirectly
CC   activates transcription in plant cells. The transgenic plants are useful
CC   for studying the absence or presence of tetracycline (Tc) or tetracycline
CC   analogues. The transgenic plants are particularly useful for identifying,
CC   comparing and characterising the activity of substances that interact
CC   with, upon or through the action of the gene product of interest. The
CC   Tc-controlled activators are useful in gene therapy for treatment of
CC   rheumatoid arthritis, hypopituitarism, tissue regeneration/wound healing
CC   cancer and AIDS. The present sequence is rabbit progesterone receptor
CC   cDNA under the control of PhCMV*-2 promoter.
XX
SQ   Sequence 4963 BP; 1259 A; 1300 C; 1278 G; 1126 T; 0 other;

  Query Match             20.6%;  Score 490.8;  DB 22;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 9e-93;
  Matches  699;  Conservative    9;  Mismatches  482;  Indels   49;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db       1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAATGACTAT 1122

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| |||| |||| || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        746 TGTGG--------- --GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
                | ||          | | | |      | |||| |     | |  | |  |
Db       1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGACATGATGGGGAGGGCAGGGGTGAA 1302

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG ---------- 843
              |  | |     | || |||  || |  ||    |||  |   ||
Db       1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                         | |    |||       ||||||   |||| ||| | ||  |  || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGCGCTTG 1422
```

```
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCATGGCCTTG 1422

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| ||| |  |  |  || |  |||||||  ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||  |  |||||||| |||| |||| |||| |||| |||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGCGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGACCTCACCCTGTTCGACCAAGTGCGCCTCTTC 1069
              ||||| |||| | | || ||||||||||  | | |||   || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAAATGCGTAGAA 1189
              |  ||||  | |||||||  |  |  | |||||||| |  |||| |||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGCAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGCAAATCTTTGACATGCTCCTGCCAACTACTTCAAGCTTTCCACAGTTAAAA 1249
              || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATCATCAAT 1782

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||    | || | | | ||  |||  | || ||  |||| |||||  |  |||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  | |    |||| | |  ||   ||     |  | |    |  | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGGGCCATCCCTGCCAGCAG 1426
              ||    | || ||  |||||  |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGTCCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || |||||||||||| || |||||  ||  | |||||||||
Db       2023 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCCGCCA 1586
              |||||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCCGCCA 2122


RESULT 40
AAI68126
ID   AAI68126 standard; cDNA; 1824 BP.
XX
AC   AAI68126;
XX
DT   13-DEC-2001  (first entry)
XX
DE   Lepomis centrarchidae oestrogen receptor encoding cDNA SEQ ID NO 5
XX
KW   oestrogen receptor; ss.
XX
OS   Lepomis centrarchidae.
XX
FH   Key             Location/Qualifiers
FT   CDS             74..1822
FT                   /*tag= a
FT                   /product= "oestrogen receptor"
XX
PN   JP2001197890-A.
XX
PD   24-JUL-2001.
XX
PF   08-NOV-2000; 2000JP-0340097.
XX
PR   09-NOV-1999;   99JP-0318113.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 2001-609402/70.
DR   P-PSDB; AAG78227.
XX
PT   New polynucleotide for controlling oestrogen receptor activity comprises
PT   the oestrogen receptor gene -
XX
PS   Claim 2; Page 25 27; 34pp; Japanese.
XX
CC   The invention relates to oestrogen receptors and the encoding genes.
CC   The oestrogen receptor genes can be used for evaluating the ability of
CC   controlling oestrogen receptor activity of a chemical substance.
XX
SQ   Sequence 1824 BP; 412 A; 554 C; 492 G; 366 T; 0 other;

  Query Match             20.5%;  Score 389.8;  DB 22;  Length 1824;
  Best Local Similarity   62.7%;  Pred. No. 1.1e-92;
  Matches  698;  Conservative    0;  Mismatches  367;  Indels   48;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||    |  ||||| ||||| ||||| || || || || || || ||||| || ||
Db        479 AACGAGATGCGTTTCTGTCTGTGTGCAGTGACTATGCCTCTGGGTACCACTACGGGGTG 538

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| |||||||| |||||||||||| ||||| || ||||| || || || ||||| |||
Db        539 TGGTCCTGTGAAGGCTGTAAGGCCTTCTTTAAGAGGAGCATCCAGGGTCACAATGACTAT 598

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCCAAGAGCTGCCAGGCC 685
              || || |||| || |||||||||| || || |  || ||| | ||||||||||||||
Db        599 ATGTGCCCAGCAACCAATCAGTGTACTATTGACAGGAATCGGCAGAAAGAGCTGCCAGGCT 658

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGACAGAGA 745
              ||||| ||| |||||||||| |||||||| || ||||  | ||    || |  ||  |
Db        659 TGCCGTCTTAGGAAGTGTTATGAACTGGGCATGATGAAAGGAGGTGTTCGCAAGGACCGT 718

Qy        746 TGTGGGTACCGCCTTGTGCGGACAGACAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGC 805
              | |  |  |  | || |  ||  | | | | |  |  |||
Db        719 GGCCGTGTTTTGCGCCGTGATAAACGACGTGCTGGAACCAATGACCGAGAGAAGGCCTCT 778

Qy        806 AAGGCCAAG----------------------AGAAGTGGCGGCCACGCGCCCCGAGTGCG 843
              |||| |  |                      || |  || || || || | |  ||
Db        779 AAGGACCTGGAGTACAAAACAGTGCCCCCTCAGGACAGGAGGAAACACAGCAGCAGCAGC 838

Qy        844 GGAGCTGCTGCTGGACG----------------CCCTGAGCCCCGAGCAGCTAGTGCTC 886
              | || | || |||| |                || ||  |    | |  ||||||
Db        839 AGTGCCGGTGGTGGAGGAGGAAAATCATCAGTGACCGGGATGTCTCCTGACCAGGTGCTC 898

Qy        887 ACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG------CCCCAGTGCGCCC 940
              ||| ||  ||| ||     |||||  |||||  |  |||       |  ||    ||
Db        899 CTCCTGCTCCAGGGTGCCGAGCCCCCAATGCTGTCCTCCCGTCACAAGCTGAGCCGACCG 958

Qy        941 TTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACAC 1000
              | |||||||| | |||| ||||     || ||||  |||||| |||||| |||| |||
Db        959 TACACCGAGGTCACCATAATGACACTACTCACCAGCATGGCCGATAAGGAGCTGGTCCAC 1018

Qy       1001 ATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTG 1060
              ||||||| |||||||||||||| |||| || ||  || ||||    ||    |||| |||
Db       1019 ATGATCACCTGGGCCAAGAAGCTTCCAGGTTTCCTGCAGCTGTCTCTGCATGACCAGGTG 1078

Qy       1061 CGCCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATT 1120
              | |||  |||||||||  ||| |||||||  | |||| |||| || || | || ||
Db       1079 CAGCTGCTGGAGAGCTCGTGGCTGGAGGTGCTGATGATTGGGCTCATATGGAGGTCCATC 1138

Qy       1121 GACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAA 1180
              ||  |||||||| |||||||| || |  || ||  | |||||||| |||| ||  |
Db       1139 CACTGCCCCGGCAAACTCATCTTCGCACAGGACCTCATACTGGACAGGAATGAAGGTGAC 1198
```

```
Qy       1181 TGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGA 1240
             || || |||||  || | || ||||| |||||||| ||||| ||| | ||  | || ||
Db       1199 TGTGTGGAAGGCTTTGTTGAGATCTTCGACATGCTGCTGGCCACTGCCTCCCGCTTCCGC 1258

Qy       1241 GAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCC 1300
               | | |||||| |||   | || | | ||||  |||| || || ||||||||||| ||
Db       1259 ATGCTCAAACTCAAACCTGAGGAGTTTGTCTGCCTCAAAGCTATCATCCTGCTCAACTCT 1318

Qy       1301 AGTATGTACCCTCTGGTCACAGCGACCCAGGATGC---TGACAGCAGCCGGAAGCTGGCT 1357
              ||   | | || |   ||| |  ||   |||  |     ||| |||| |    ||
Db       1319 GGTGCCTTCTCTTTCTGCACCGGCACAATGGAGCCCCTCCACAACAGCATGGCAGTGCAG 1378

Qy       1358 CACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCC 1417
              || ||||| || || . || || ||| |  |        ||   | |    ||   |||
Db       1379 AACATGCTGGACACCATCACAGACGCTCTCATACATCATATCAGCCAATCAGGATGCTCG 1438

Qy       1418 TCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCAT 1477
              | |||||||| || | ||| | |||  | || ||| | || || |||||| |||||||
Db       1439 GCTCAGCAGCAGTCGAGGCGGCAGGCCCAGCTGCTGCTCCTGCTCTCCCACATCAGGCAC 1498

Qy       1478 GCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTG 1537
               ||| ||||| |||||||| |||||   || ||||||||||| ||    || ||  |
Db       1499 ATGAGCAACAAAGGCATGGAGCATCTCTACAGCATGAAGTGCAAGAACAAAGTGCCTCTT 1558

Qy       1538 TATGACCTGCTGCTGGAGATGCTGAATGCCCAC 1570
              || ||||| |||||||||.|||| || | || |||
Db       1559 TACGACCTTCTGCTGGAGATGTTGGACGCTCAC 1591


Search completed: June 16, 2003, 00:48:03
Job time : 441.175 secs
```

Geneseq version 5.1.2
Copyright (c) 1989 - 2003 Compugen Ltd.

Matrix: nucleic search, using sw model

Run on: June 15, 2003, 21:19:01 ; Search time 428.760 Seconds
(without alignments)
9822.[illegible] Million cell updates/sec

Title: US-08-826-361A-1
Perfect score: 1434
Sequence: 1 ATGGAATTATAGGGGTAT...GGAAAGGGG

83	203	14.2	606	21	AR027646	Human, genome seq
84	203	14.2	597	22	ABA61882	Human, foetal liver
85	203	14.2	597	22	ABA29437	Probe #7908 for qe
86	203	14.2	597	22	AAK10196	Human brain expres
87	203	14.2	597	22	AAK86003	Human, bone marrow
88	203	14.2	597	22	AA141808	Probe #10494 used
89	203	14.2	597	24	AR010245	Human genome deriv
90	189	13.2	599	22	ABA59743	Human foetal liver
91	189	13.2	599	22	ABA28652	Probe #6718 for qe
92	189	13.2	599	22	AAK08012	Human brain expres
93	189	13.2	599	22	AAK83884	Human bone marrow
94	189	13.2	599	22	AA139610	Probe #8296 used f
95	189	13.2	599	24	AR008715	Human genome deriv
96	188	12.8	188	22	ABA73285	Human foetal liver
97	188	12.8	188	22	ABA40142	Probe #16598 for q
98	188	12.8	188	22	AAK20797	Human brain expres
99	188	12.8	188	22	AAK16853	Human bone marrow
100	188	12.8	188	22	AA152692	Probe #21478 used
101	188	12.8	188	24	AR021150	Human genome deriv
102	157.6	11.0	569	20	AAX37613	pDM16 oestrogen re
103	157	10.9	521	20	AAX37808	pDM12 oestrogen re
104	157	10.9	566	20	AAX37812	pDM14 oestrogen re
105	153.8	10.7	1470	22	AA306086	Angiotensin conver
106	153.8	10.7	1560	22	AA306158	Angiotensin conver
107	142.2	9.9	284	19	AAU85976	Human, oestrogen re
108	138.8	9.6	2218	24	AR064869	Human cDNA differe
109	138.8	9.6	2218	24	AR064900	Stomach cancer rel
110	129	9.0	2249	9	AAN80921	Sequence encoding
111	129	9.0	2569	22	AA158567	Human polynucleoti
112	129	9.0	4,250	21	ABA62902	Human, nuclear hor
113	121.2	8.5	2559	22	AA160454	Human polynucleoti
114	118.8	8.3	1377	21	AAA96726	Human ERRgamma cod
115	118.8	8.3	1441	21	AAZ95741	Human oestrogen re
116	118.8	8.3	1615	19	AAV12645	steroid hormone re
117	118.8	8.3	2985	20	AAX42266	Human nuclear rece
118	118.8	8.3	2985	21	AAA09801	Human nuclear rece
119	118.8	8.3	2987	20	AAX42267	Human nuclear rece
120	118.8	8.3	2987	21	AAA09802	Human nuclear rece

ALIGNMENTS

```
RESULT 1
AA188412
ID   AA188412 standard; DNA; 31 BP.
XX
AC   AA188412;
XX
DT   29-APR-1998 (first entry)
XX
DE   Human oestrogen receptor protein cDNA.
XX
KW   oestrogen receptor protein; steroid; alternative splicing; oestradiol;
KW   estrone; estriol; screening; ds.
XX
OS   Homo sapiens.
XX
RN   [1]
RA   Green S.;
XX
RT   ;
XX
RL   Submitted (29-APR-1997) to the EMBL/GenBank/DDBJ databases.
XX
RL   Updated (15-MAR-1998) to the EMBL/GenBank/DDBJ databases.
RL   Updated (20-NOV-1998) to the EMBL/GenBank/DDBJ databases.
RL   Updated (29-MAR-1999) to the EMBL/GenBank/DDBJ databases.
XX
RA   (AVEX ) AVEX NOBEL NV
XX
RT   Klokova E.; Moskvitin S.;
XX
RP   WO9: 1997-US10844.
RP   P-ISSUE: AAW43212.
XX
PI   DNA encoding oestrogen receptor - useful in screening assays
PI   identify novel ligands - hormonal analogues
XX
PS   Claim 1; Page 14; 19pp; English.
XX
CC   This sequence encodes a novel oestrogen binding protein. The
CC   sequence which encodes this protein can be alternatively spliced
CC   resulting in the detection of additional transcripts (see AA188413).
CC   This receptor is able to bind and be activated by estradiol, estrone and
CC   estriol, can be used in a screening assay for the identification of
CC   drugs which mimic hormonal analogues.
XX
SQ   Sequence 31 BP; 11 A; 5 C; 4 G; 0 T; 0 other;

  Query Match          Score 1434; DB 18; Length 31;
  Best Local Similarity  Pred. No. 0;
  Matches 31; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

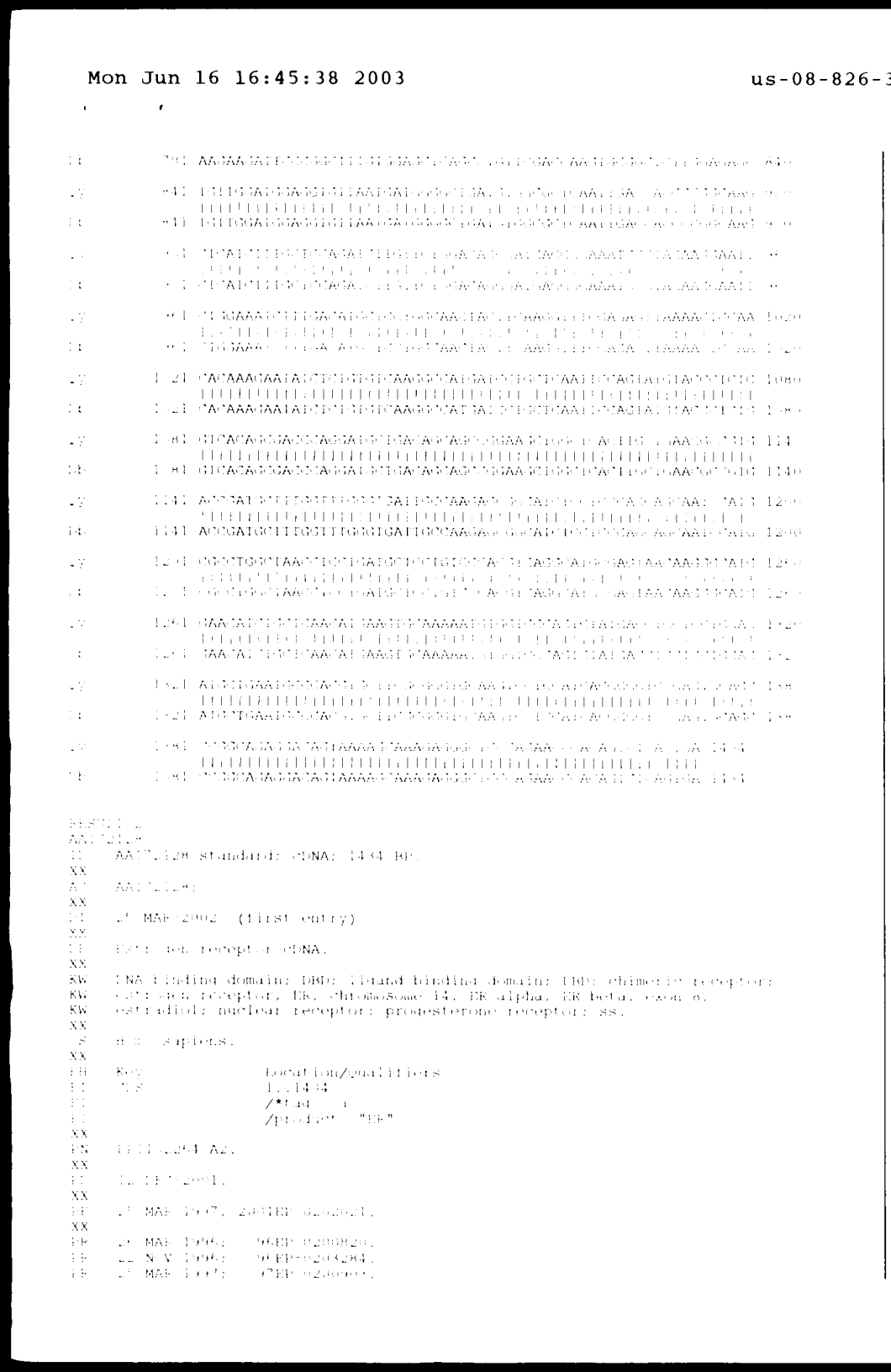
Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

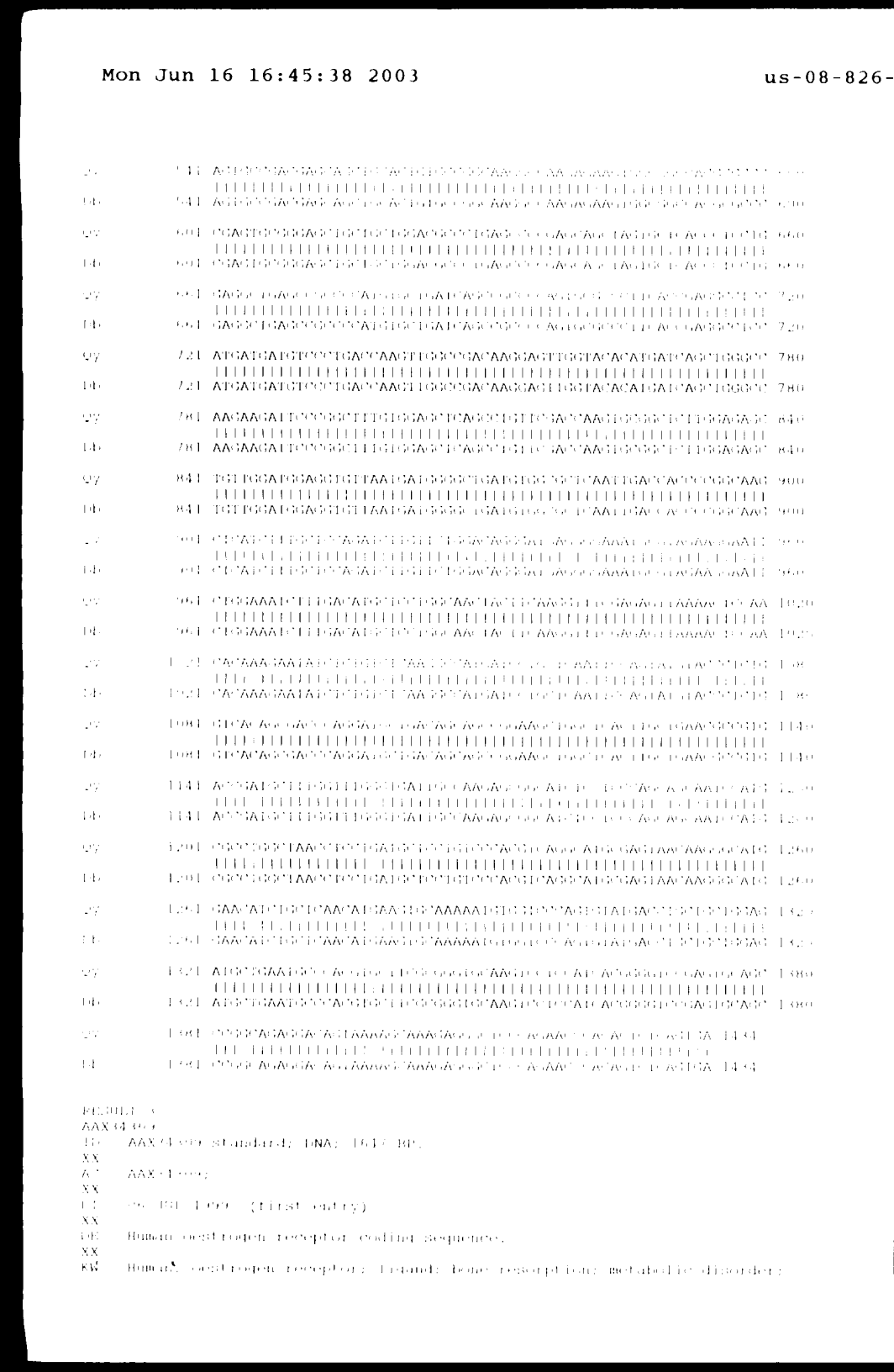
Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31

Qy          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
              ||||||||||||||||||||||||||||||| |
Db          1 AAAAAAAAAAACCCCAAGGGGGGGGGGTTTTT 31
```

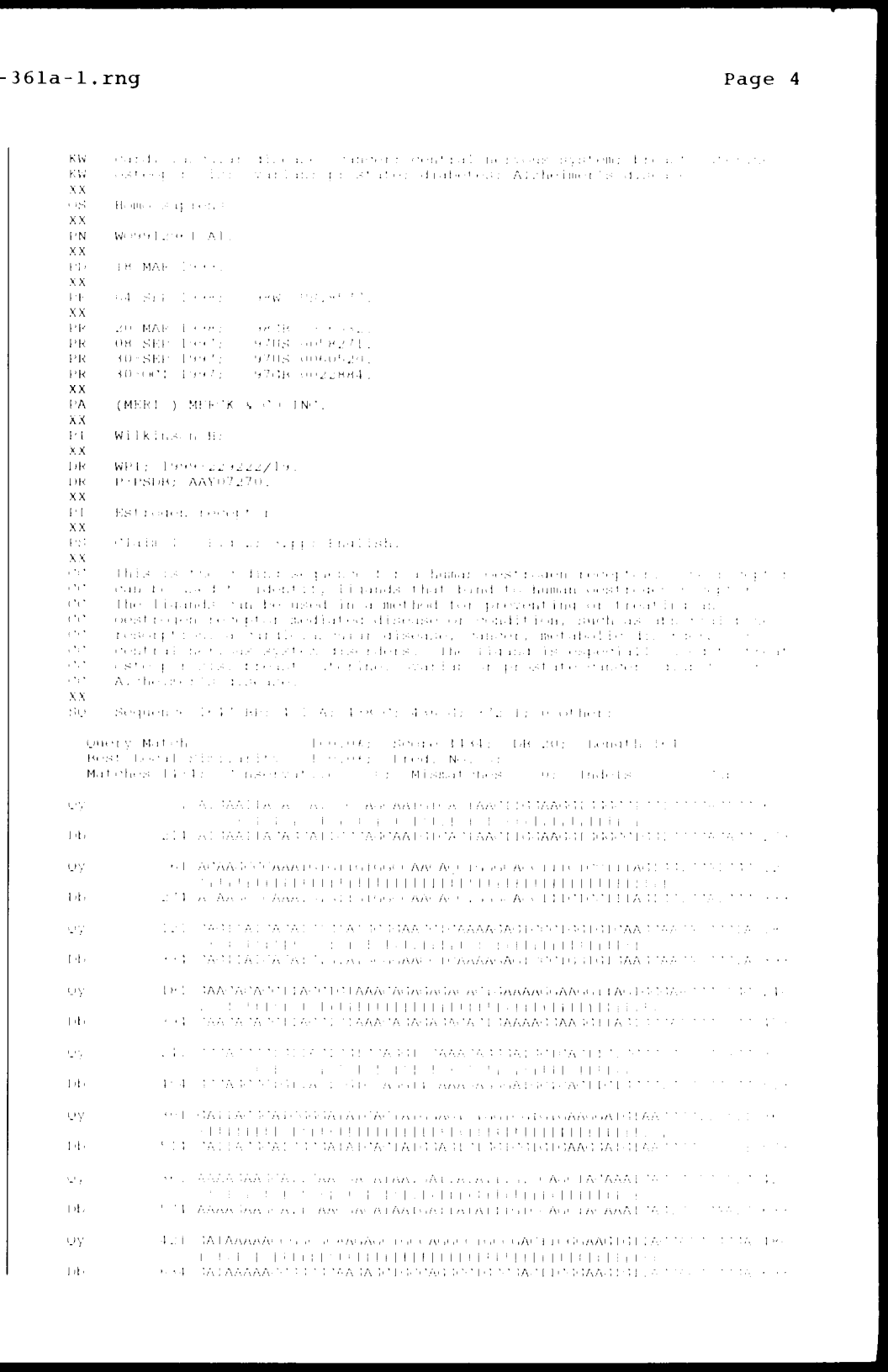


XX
PA (ALKU) AKC N BEL NV.
XX
RT Mossorba S. [Kenn E]
XX
DR WPI; 2[redacted]/12.
DR [redacted]AAS[redacted].
XX
PT New [redacted] binding to [redacted] expressed DNA binding
PT ligand binding domain of a new estrogen receptor, for [redacted]
PT [redacted] breast cancer [redacted] for the receptor.
XX
PS Example A; [redacted] pages; [redacted].
XX
CC This sequence encodes a novel estrogen receptor (ER). The [redacted]
CC this new ER is located on chromosome 14 and has a different [redacted]
CC distribution to [redacted] ER. This ER also has [redacted]
CC ERalpha and is [redacted]. These orphan receptors have [redacted]
CC related structure but do not appear to be able to bind [redacted]
CC ER ligands. The DNA binding domain (DBD) and ligand binding [redacted] (LBD)
CC from this ER may be used in the chimeric receptor of the [redacted]
CC which also has an N terminal domain. The chimeric receptor [redacted]
CC encoding it, be used in a screening assay to identify [redacted]
CC drugs. Similar chimeric receptors comprising the LBD of [redacted]
CC and also comprising the LBD and an N-terminal domain derived [redacted]
CC another nuclear receptor e.g. progesterone receptor, [redacted]
CC screening [redacted] may be used to identify new ligands or [redacted]
CC are able to [redacted] to new [redacted] chimeric receptors comprising [redacted]
CC of the new ER, and LBD of [redacted] N terminal domain derived [redacted]
CC another nuclear receptor, [redacted] be used to identify new [redacted]
CC hormone analogs for the nuclear receptors.
XX
SQ Sequence [redacted] AA; [redacted] A [redacted] C [redacted] G [redacted] T [redacted] other;

Query Match [redacted] Score [redacted]; DB [redacted]; Length [redacted];
Best Local Similarity [redacted]; Pred. No. [redacted];
Matches [redacted]; Conservative [redacted]; Mismatches [redacted]; Indels [redacted]; Gaps [redacted];

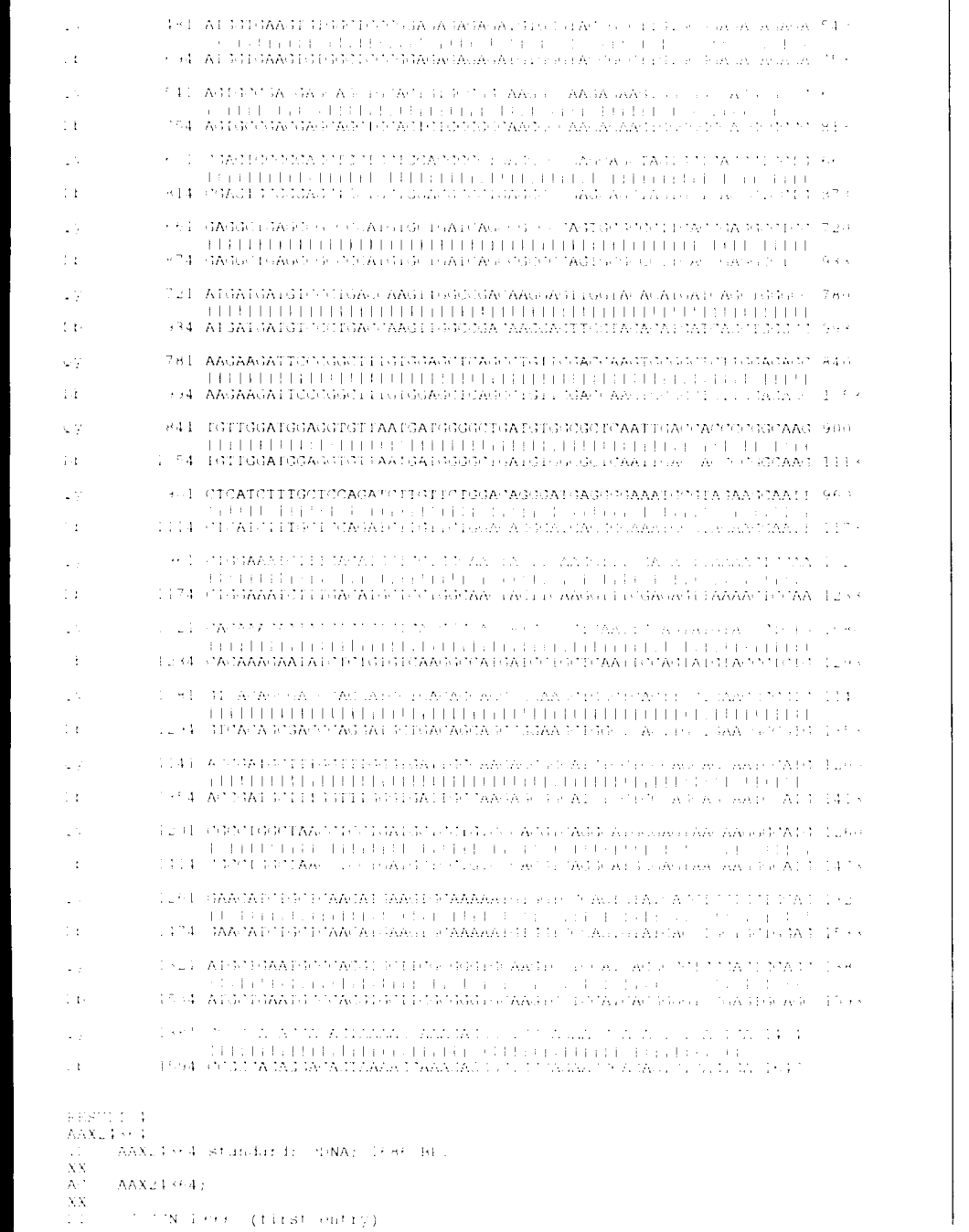


RESULT 1
AAX84869
ID AAX84869 standard; DNA; 1617 BP.
XX
AC AAX84869;
XX
DT 09-SEP-1999 (first entry)
XX
DE Human oestrogen receptor coding sequence.
XX
KW Human; oestrogen receptor; ligand; bone resorption; metabolic disorder;



KW cardiovascular disease; infertility; contraception; hypertension; breast cancer;
KW osteoporosis; CNS; Parkinson's disease; Alzheimer's disease;
XX
OS Homo sapiens.
XX
PN WO9912966-A1.
XX
PD 18-MAR-1999.
XX
PF 04-SEP-1998; 98WO-US018411.
XX
PR 05-SEP-1997; 97US-0057871.
PR 08-SEP-1997; 97US-0058201.
PR 30-OCT-1997; 97US-0064884.
XX
PA (MERI) MERCK & CO INC.
XX
PI Wilkins S B;
XX
DR WPI; 1999-224222/19.
DR P-PSDB; AAY07270.
XX
PT Estrogen receptor 1
XX
PS Claim 1; Page 22; 57pp; English.
XX
CC This is the human oestrogen receptor (ER) coding sequence
CC cited in the claims. Ligands that bind to human oestrogen receptor
CC the ligands can be used in a method for preventing or treating a
CC oestrogen receptor mediated disease or condition, such as breast cancer
CC receptor positive cancer, osteoporosis, obesity, metabolic bone disorders
CC cardiovascular disorders, hypertension, infertility, contraception,
CC CNS disorders, Parkinson's disease, Alzheimer's disease, hot flashes,
CC A. thaliana cDNA sequence.
XX
SQ Sequence 1617 BP; 411 A; 399 C; 464 G; 342 T; 0 other;

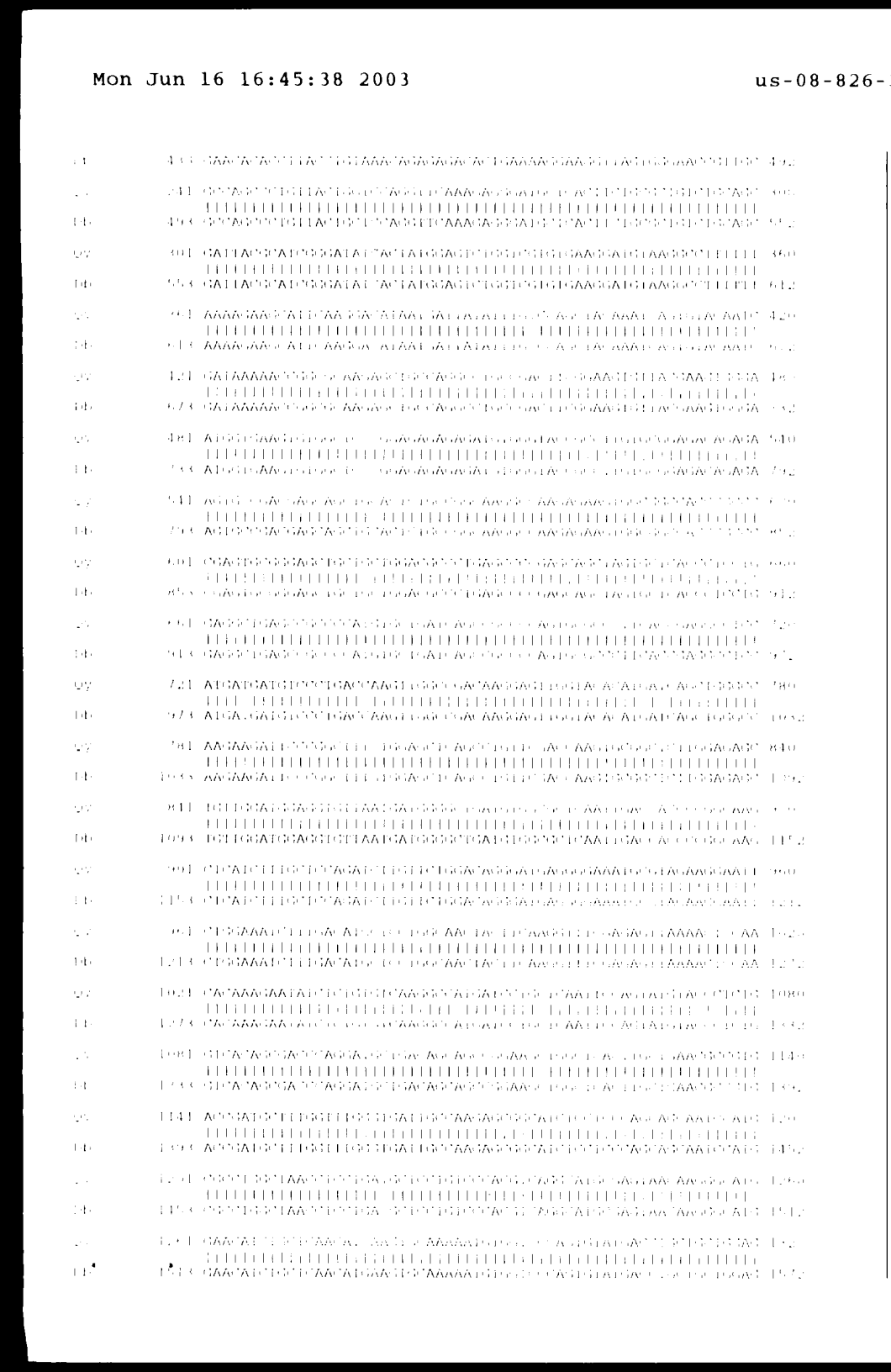
Query Match 100.0%; Score 1161; DB 20; Length 1617
Best Local Similarity 100.0%; Pred. No. ;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

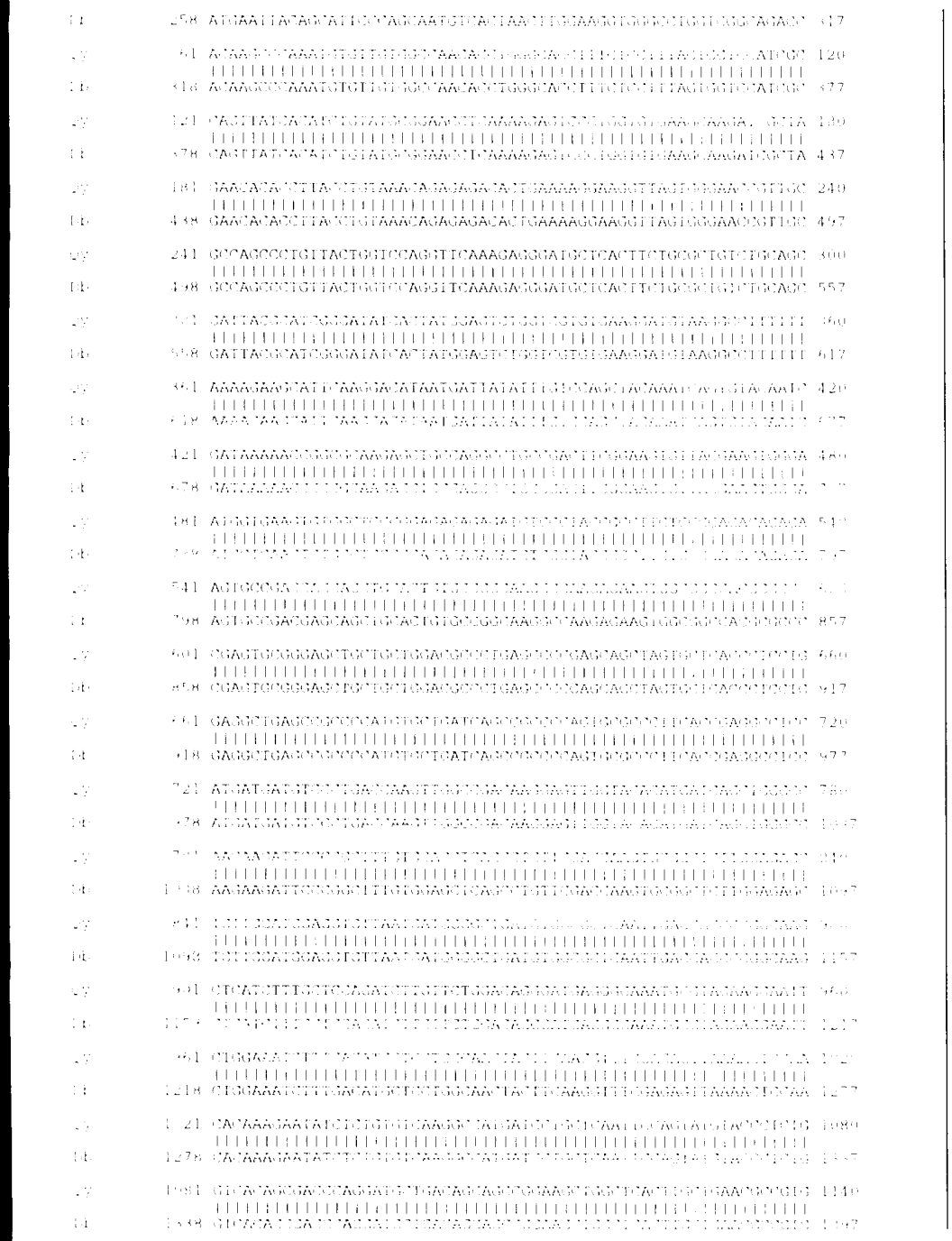


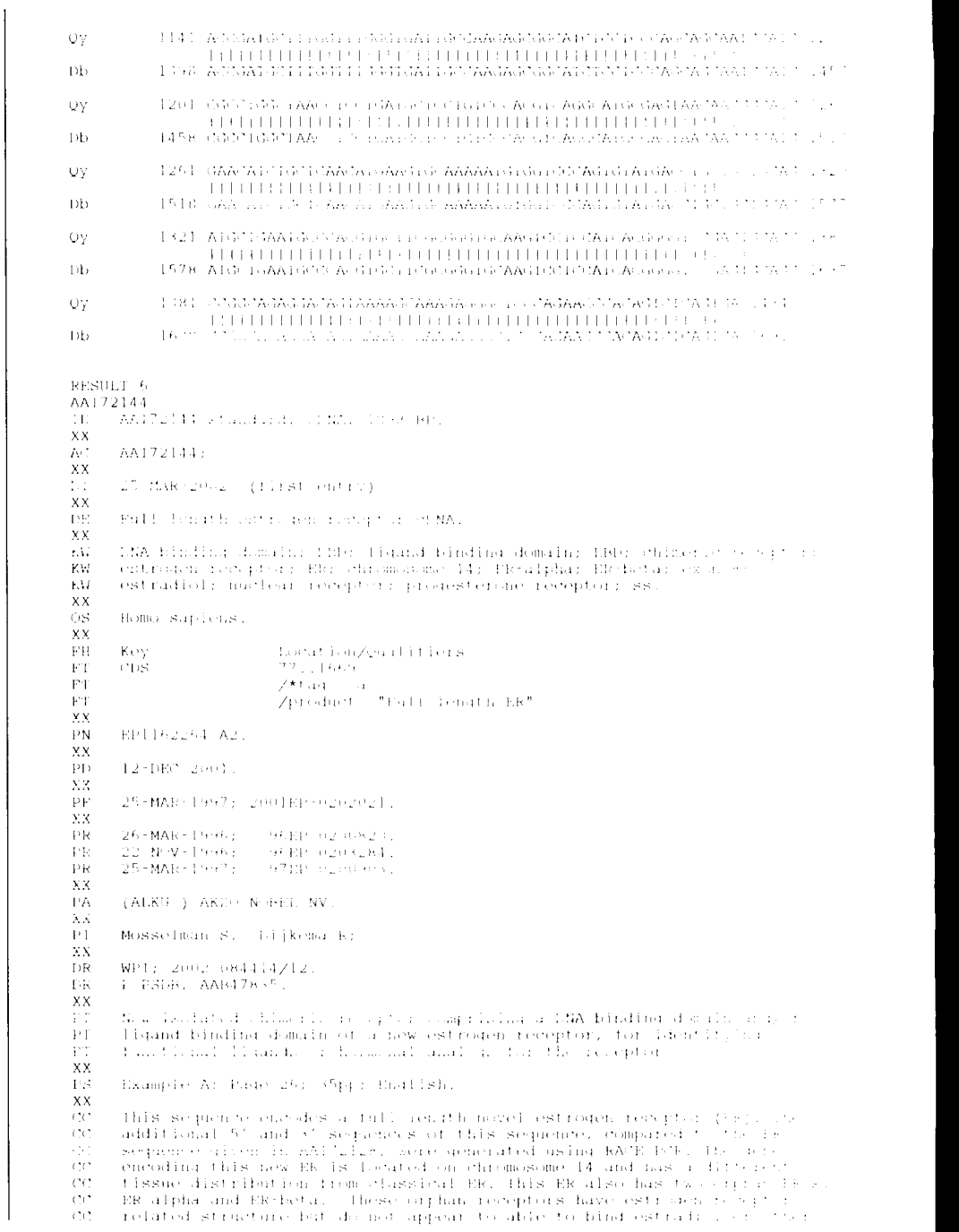
RESULT 1
AAX21964
ID AAX21964 standard; DNA; 2486 BP.
XX
AC AAX21964;
XX
DT 16-JUN-2003 (first entry)

XX
DE Human oestrogen receptor beta cDNA.
XX
KW oestrogen receptor beta binder; hormone; ss
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS
FT /*tag=
XX
PN WO0007847-A1.
XX
PD 17-FEB-2000.
XX
PF 29-JUL-1999; 99WO-US017431.
XX
PR 05-AUG-1998; 98US-0095564P.
XX
PA (AMBP) AMERICAN H MESSAGE PROD.
XX
PI Bhat RA, Hershberger PA, Nettles KW, Katzenellenbogen JA;
XX
DR WPI; 2000-217121/18.
DR P-PSDB; AAW07464.
XX
PT Human oestrogen receptor beta coding sequence - useful in the
PT production of human oestrogen receptor beta and identifying novel
PT human oestrogen receptor beta interactive compounds
XX
PS Claim 1; Page 11; 62pp; English.
XX
CC This cDNA encodes a full length human oestrogen receptor beta
CC (hER beta, see AAW07464). The receptor is a previously cloned member
CC of the steroid receptor superfamily of transcription factors. Human
CC cDNA was isolated from a PCR library prepared from MCF-7 cells
CC supplemented with breast, ovarian and prostate human sequences which
CC missing N terminal residues (Met and Thr) found in rat ER beta
CC and to supply an additional Kozak translation initiation sequence.
CC A truncated splice variant of hER beta was also cloned and
CC its cDNA is described. The PCR sequence was generated
CC by PCR amplification from a human ovary cDNA library and hER beta
CC (see AAX21963). ER beta is also highly expressed in the
CC spleen, testis and thymus. The clones encompass the hER beta
CC polymorphisms and polymorphisms, particularly peptides which
CC include residues 247 and 255 of hER beta. The invention also
CC expression systems (e.g. which transform, transfect or infect
CC transformed cells) can be produced, as well as screening methods to
CC identify/use hER beta agonists and antagonists (including
CC tissue specific agonists and antagonists) and hER beta
CC co-activator proteins.
XX
SQ Sequence 2486 BP; 611 A; 715 C; 611 G; 549 T; 0 other;

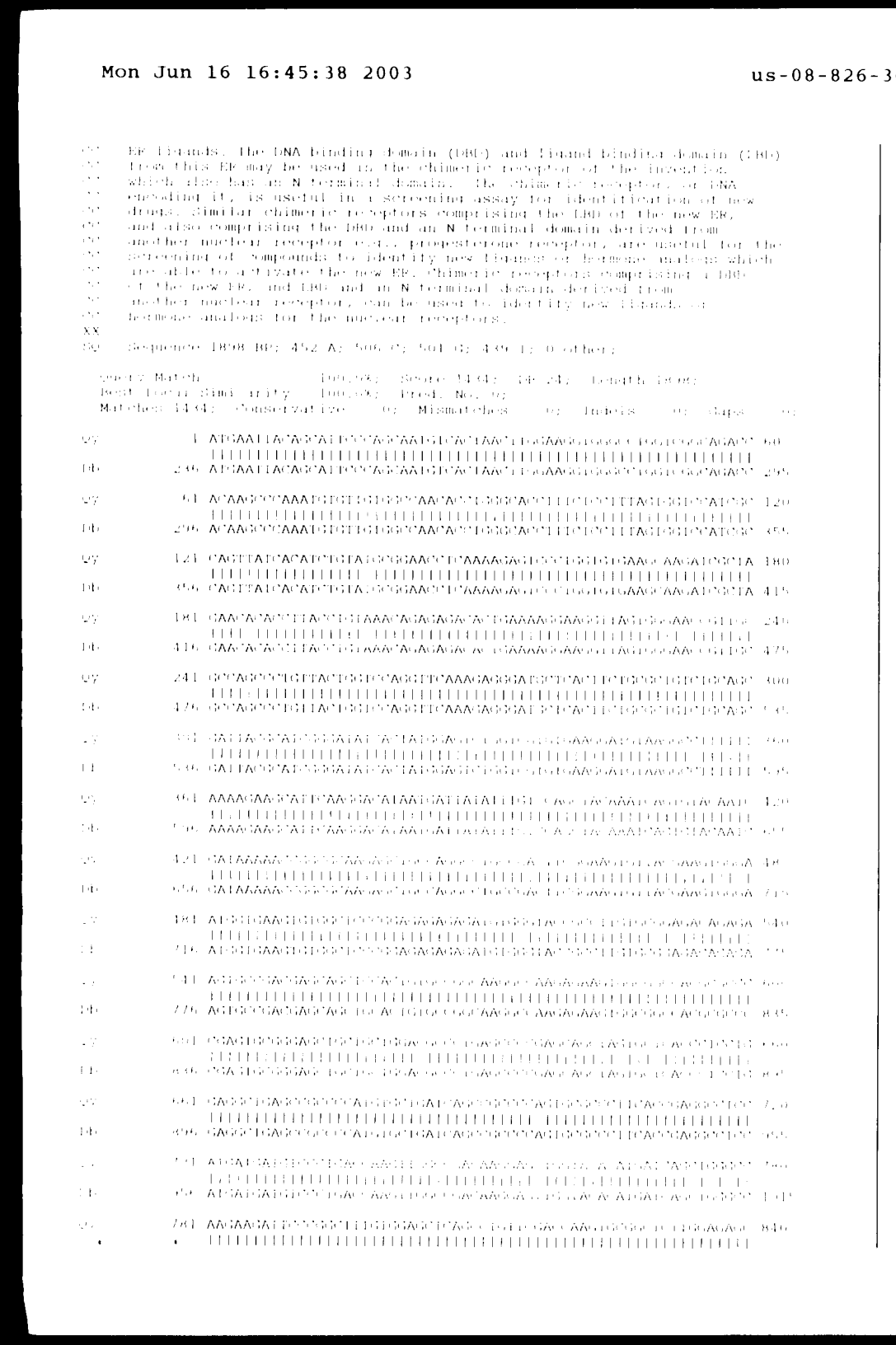
Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible]

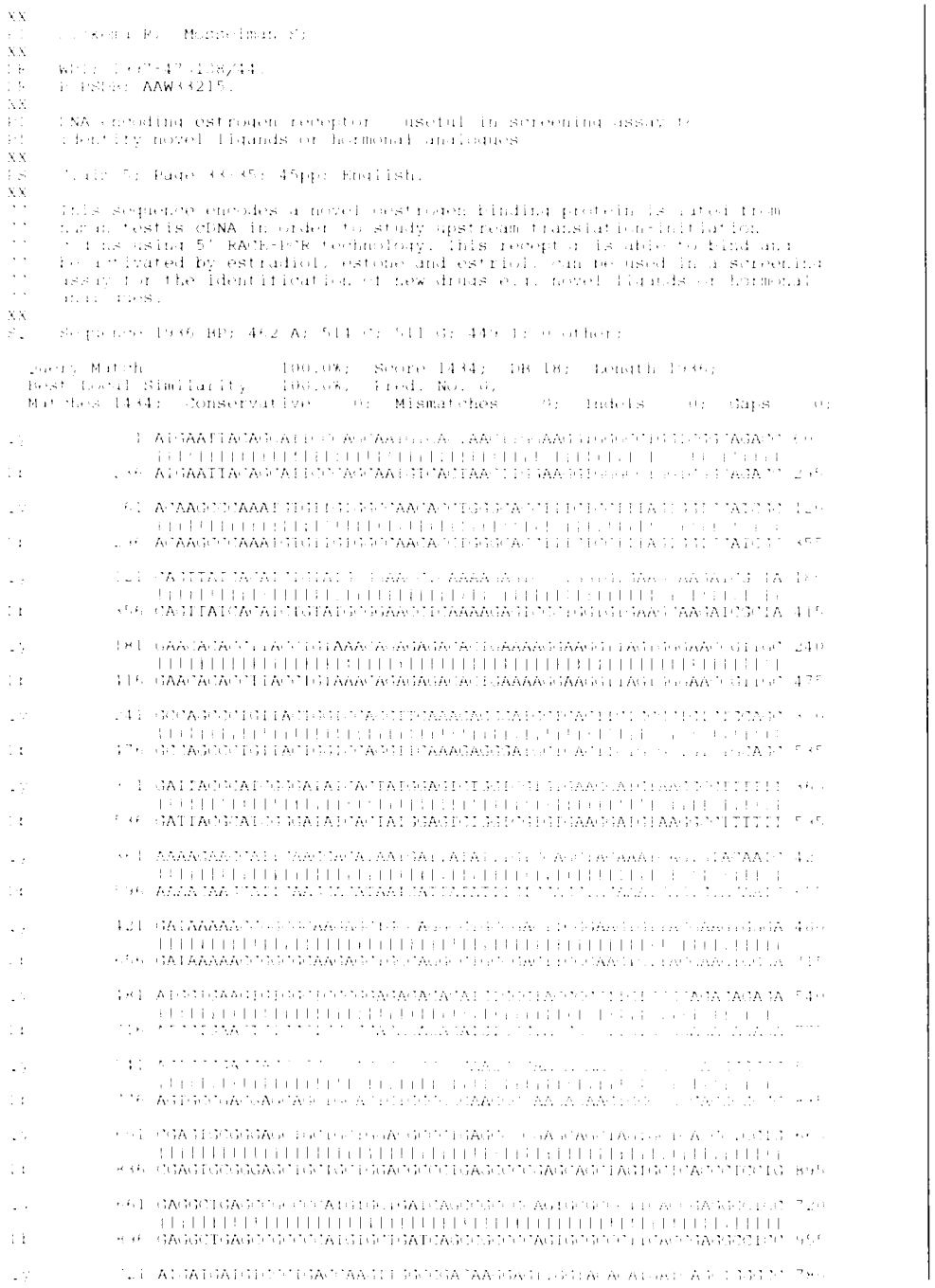


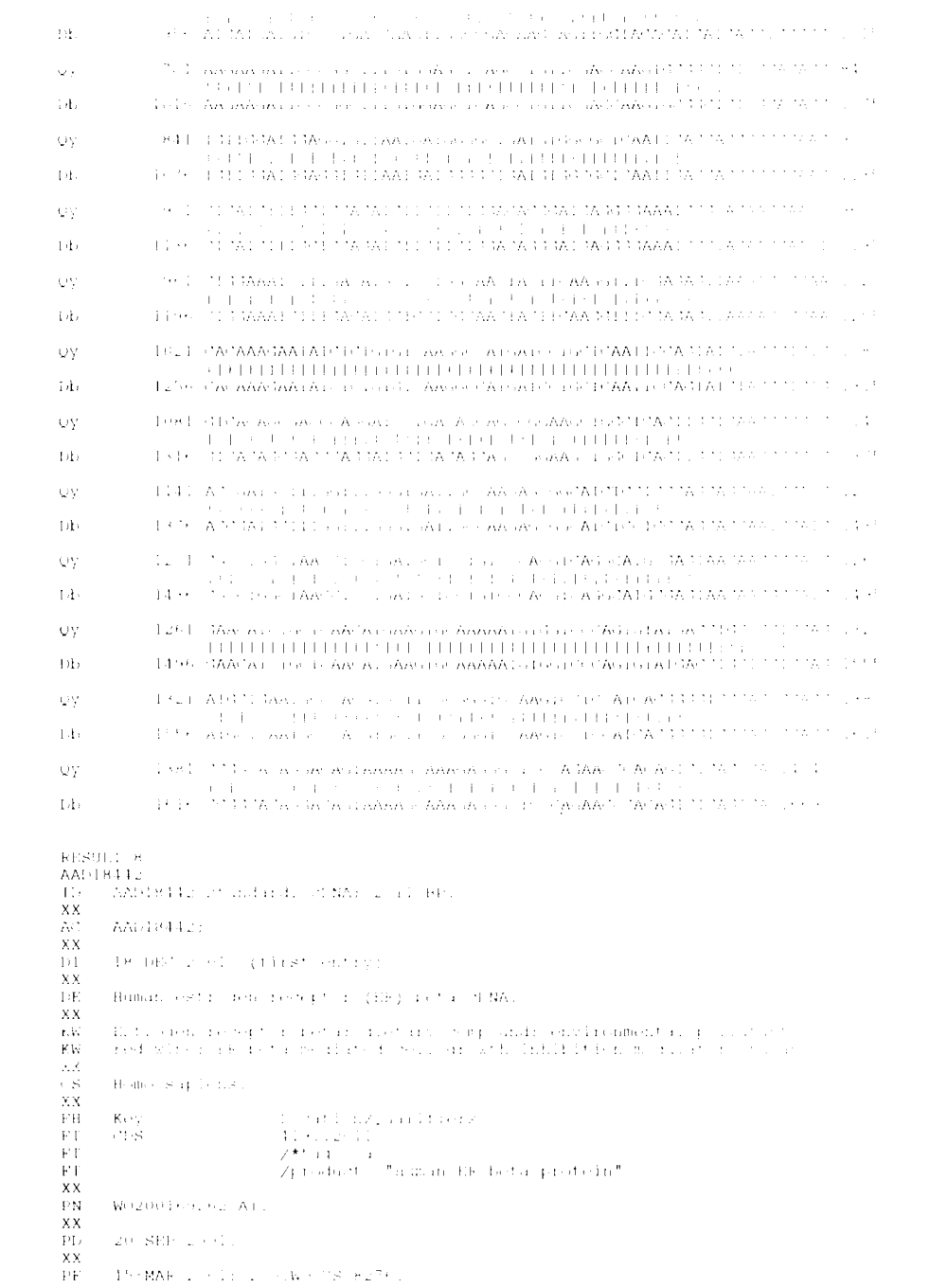




RESULT 6
AA172144
ID AA172144 standard; DNA; 1744 BP.
XX
AC AA172144;
XX
DT 27-MAR-2002 (first entry)
XX
DE Full length cDNA sequence.
XX
KW cDNA Binding domain; LBD; ligand binding domain; LBD; chimeric protein;
KW estrogen receptor; ER; chromosome 14; FR-alpha; ER-beta;
KW ostradiol; nuclear receptor; prostostone receptor; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 77..1699
FT /*tag= a
FT /product= "Full length ER"
XX
PN EP1162264 A2.
XX
PD 12-DEC-2001.
XX
PF 25-MAR-1997; 2001EP-0202021.
XX
PR 26-MAR-1996; 96EP-0200824.
PR 22-NOV-1996; 96EP-0203284.
PR 25-MAR-1997; 97EP-0200855.
XX
PA (AKZO) AKZO NOBEL NV.
XX
PI Mosselman S, Dijkema R;
XX
DR WPI; 2002-084114/12.
DR P-PSDB; AA847855.
XX
PT New isolated human estrogen receptor, useful in DNA binding domain
PT ligand binding domain of a new estrogen receptor, for identifying
PT functional fragments of hormonal agonist to the receptor
XX
PS Example A; Page 26; 56pp; English.
XX
CC This sequence encodes a full length novel estrogen receptor (ER). The
CC additional 5' and 3' sequences of this sequence, compared to the
CC sequence given in AAI71125, were generated using RACE PCR. The gene
CC encoding this new ER is located on chromosome 14 and has a different
CC tissue distribution from classical ER. This ER also has two splice
CC ER alpha and ERbeta. These orphan receptors have ostrogen receptor
CC related structure but do not appear to able to bind ostradiol and







```
XX
RP   1; MAR 2000; 2000WO-IB000051.
XX
RA   (UYMA ) UNIV MASSACHUSETTS.
XX
PT   Bo S;
XX
DR   WPI; 2001-500920/67.
DR   P-PSDB; AAE10457.
XX
PI   In vitro screening for modulator of estrogen receptor beta mediated
PI   cell growth inhibition by contacting cell having functional ER beta
PI   protein with candidate compound and detecting ER beta regulated gene
PI   expression
XX
DS   Example 3; Fig 7B; Pages; English.
XX
CC   The invention relates to in vitro screening for modulator of estrogen
CC   receptor beta (ER beta) mediated cell growth inhibition. The method
CC   involves contacting mammalian cell having functional ER beta protein
CC   with candidate compound and detecting increase/decrease of ER-beta
CC   regulated gene expression in presence of candidate compound compared
CC   to expression in absence of the compound. The method is useful for
CC   identifying a compound that modulates ER beta mediated cell growth
CC   inhibition. It is useful for testing dietary compounds, e.g. red wine,
CC   for the presence of estrogen mimics that trigger ER beta mediated
CC   processes, thus affecting physiological function positively or
CC   negatively and for testing environmental pollutants for the presence
CC   of estrogen mimics that may pose health risks involving ER beta
CC   mediated processes. The present sequence is human ER beta cDNA.
XX
SQ   Sequence 2011 BP; 470 A; 54 C; 561 G; 466 T; 0 other;

  Query Match             100.0%;  Score 1434;  DB 22;  Length 2011;
  Best Local Similarity   100.0%;  Pred. No. 0.;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGTATTCCAGTAATGCACTAATTCCAATCGGCTTCCCGGTGAAT 54
              ||||||
```

KW ...immune disease; systemic lupus erythematosus; arthritis; rheumatism;
KW osteoarthritis; osteoporosis; breast cancer; endometrial cancer.
XX
OS Homo sapiens.
XX
PN WO200162793-A2.
XX
PD ...-AUG-2001.
XX
PF ... FEB 2001; 2001WO-US05860.
XX
PR 22-FEB-2000; 2000US-0184755.
PR 21-JAN-2001; 2001US-0764147.
XX
PA (UEAB) UNIV ABERDEEN.
XX
PI Kuiper E, Cassel MJ, Black DS, Wink Jeen HC.
XX
DR WPI; 2001-532011/57.
DR P-PSDB; AAU27322.
XX
PT ...ast and receptor gene and protein polymorphisms useful for diagnosis
PT of individuals at risk of developing bone disorders -
XX
PS Disclosure; Figure 4; 245pp; English.
XX
CC The invention relates to a novel isolated peptide comprising or
CC consisting of an amino acid sequence selected from an amino acid sequence
CC of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
CC ... amino acids), antibodies against them, nucleic acids encoding
CC them (including vectors for transforming cells), the gene for human
CC ERbeta is located on chromosome 6q.25.1. The variants are encoded
CC by single nucleotide polymorphisms (SNP). The variant peptides and
CC proteins can be used in assays to determine the biological
CC a...ity of the protein, to raise antibodies, as a reagent in assays
CC designed to quantitatively determine levels of the protein in
CC biological fluids, to identify compounds that modulate receptor
CC activity and to screen compounds for the ability to stimulate or
CC inhibit interaction between the receptor protein and a target molecule
CC that normally interacts with the receptor protein e.g. oestrogen.
CC The antibody can be used to isolate the protein, to assess expression in
CC disease states e.g. cardiovascular disease and autoimmune disease (e.g.
CC systemic lupus erythematosus, arthritis, rheumatism and osteoarthritis),
CC osteoporosis, breast cancer and endometrial cancer. In addition
CC the antibodies can be used in pharmacokinetic analysis and inhibiting
CC protein function, e.g. blocking the binding of the oestrogen receptor
CC protein to a binding partner such as a ligand. The nucleic acids
CC encoding the proteins can be used as probes, primers, chemical
CC ...ntermediates in PCR assays. The present sequence is a ...
CC human ERbeta.
XX
SQ Sequence 2011 BP; 470 A; 541 C; 561 G; 439 T; 0 other;

Query Match 100.0%; Score 1434; DB 22; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0.;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAATTACAACATCCCAGAAATTCTACTAACTTCCTGAAGTCGGCAGTCGTCATGAC 60
Db ... ATGGAATTACAGCATCCCAGAAATTCTACTAACTTCCTGAAGTCGGCAGTCGTCATGAC 647

Qy 61 ATGAGCCTAAATCTGCTCGATGGTGTGAACTCTGCCATTCTCTCCTTTAGTTCTATCCAG 120
Db ... ATGAGCCTAAATCTGCTCGATGGTGTGAACTCTGCCATTCTCTCCTTTAGTTCTATCCAG 707

Qy 121 CAGTTATCAGTAGTCTTAATGGTGAACCTAAAATAATTCCTCACATCGATATCGTA 180
Db ... CAGTTATCAGTAGTCTTAATGGTGAACCTAAAATAATTCCTCACATCGATATCGTA 767

Qy 181 GAACATACATTTACTGTTAAAGAGAAATATGTGAAAATGAAATTAGTATAAATCTTGCC 240
Db ... GAACATACATTTACTGTTAAAGAGAAATATGTGAAAATGAAATTAGTATAAATCTTGCC 827

Qy 241 GCTATGCCTGTTACTGGTCCAGTTCAAAGAGGATGCTCACTTCTTTTTTTTTTTTTTT 300
Db ... GCTATGCCTGTTACTGGTCCAGTTCAAAGAGGATGCTCACTTCTTTTTTTTTTTTTTT 887

Qy 301 TATTATCCATCTGAATATTATTATCTATCTCCTCCTGTTTCAAGTATTTTTTTTTTTTTT 360
Db ... TATTATCCATCTGAATATTATTATCTATCTCCTCCTGTTTCAAGTATTTTTTTTTTTTTT 947

Qy 361 TT... 420
Db ... TT... 1007

Qy 421 TT... 480
Db ... TT... 1067

Qy 481 TT... 540
Db ... TT... 1127

Qy 541 TT... 600
Db ... TT... 1187

Qy 601 TTAGTTGCTAATCTTCTTGAGTTAGATGATAGAGATCTAGTTCATTTTTTTTTTTTTT 660
Db ... TTAGTTGCTAATCTTCTTGAGTTAGATGATAGAGATCTAGTTCATTTTTTTTTTTTTT 1247

Qy 661 TAGGGTGATGGTTCTTTTACTCCTGAAGAGACTGCTCAGTTGGGTCAAGATGGTTTTTTT 720
Db ... TAGGGTGATGGTTCTTTTACTCCTGAAGAGACTGCTCAGTTGGGTCAAGATGGTTTTTTT 1307

Qy 721 ATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 780
Db ... ATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1367

Qy 781 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 840
Db ... TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1427

Qy 841 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 900
Db ... TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1487

Qy 901 CTCTATCTTTTGCTGCTAGATCTTGTTAATGCTAGGGAAAATCTTGATAAATCTTTAT 960
Db ... CTCTATCTTTTGCTGCTAGATCTTGTTAATGCTAGGGAAAATCTTGATAAATCTTTAT 1547

Qy 961 TT... 1020
Db ... TT... 1607

Qy 1021 CATAAAGAATATCTTTTTTTTAATGCTATGATTCCTCCTGCTAATTTCCCTAATTTCTAT 1080
Db ... CATAAAGAATATCTTTTTTTTAATGCTATGATTCCTCCTGCTAATTTCCCTAATTTCTAT 1667

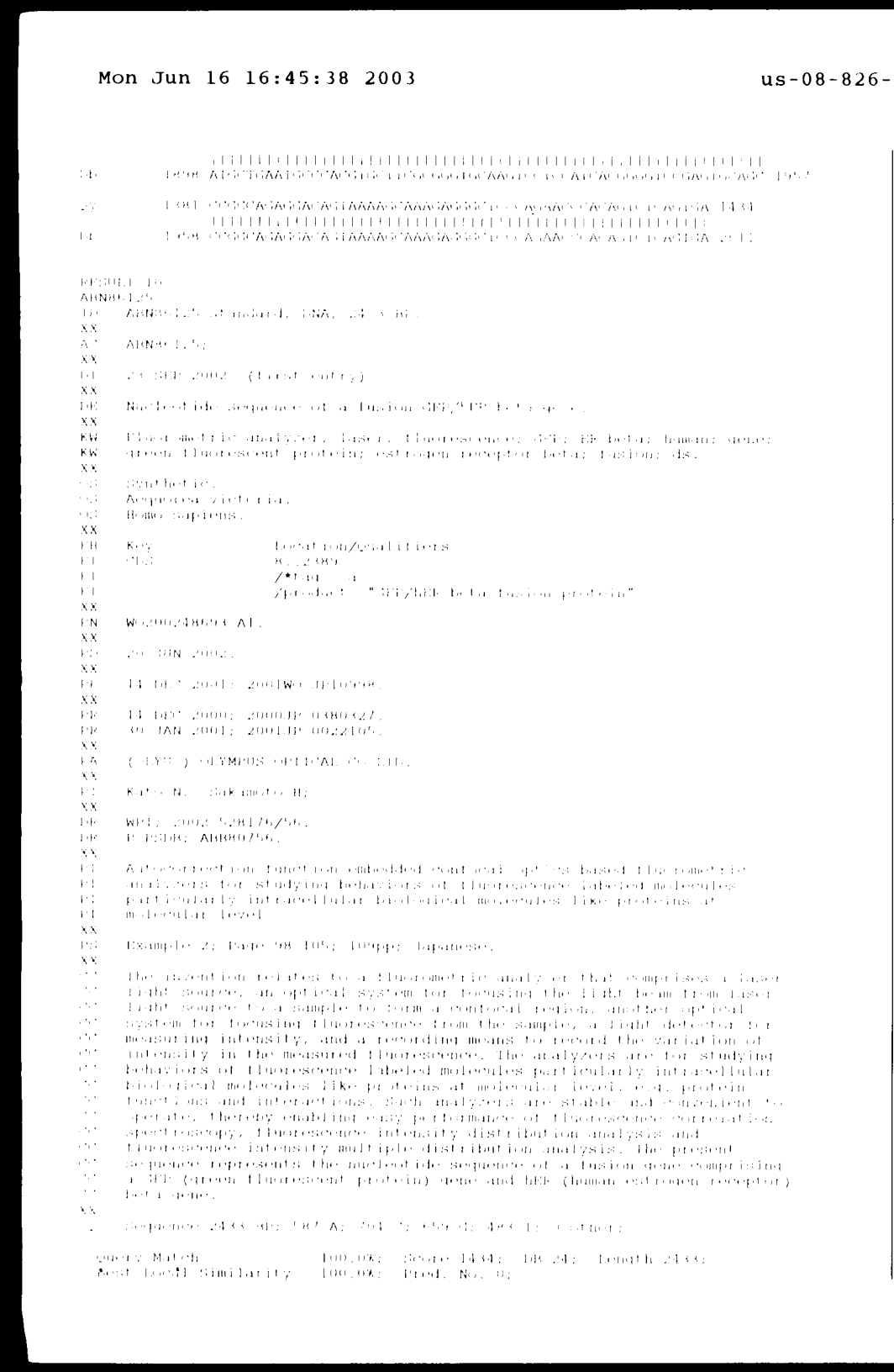
Qy 1081 GTTCCTACTCCTGAATATTCTATTCAAGATGCTGAAGGTTTTGAAATTGAATTTATTGAT 1140
Db ... GTTCCTACTCCTGAATATTCTATTCAAGATGCTGAAGGTTTTGAAATTGAATTTATTGAT 1727

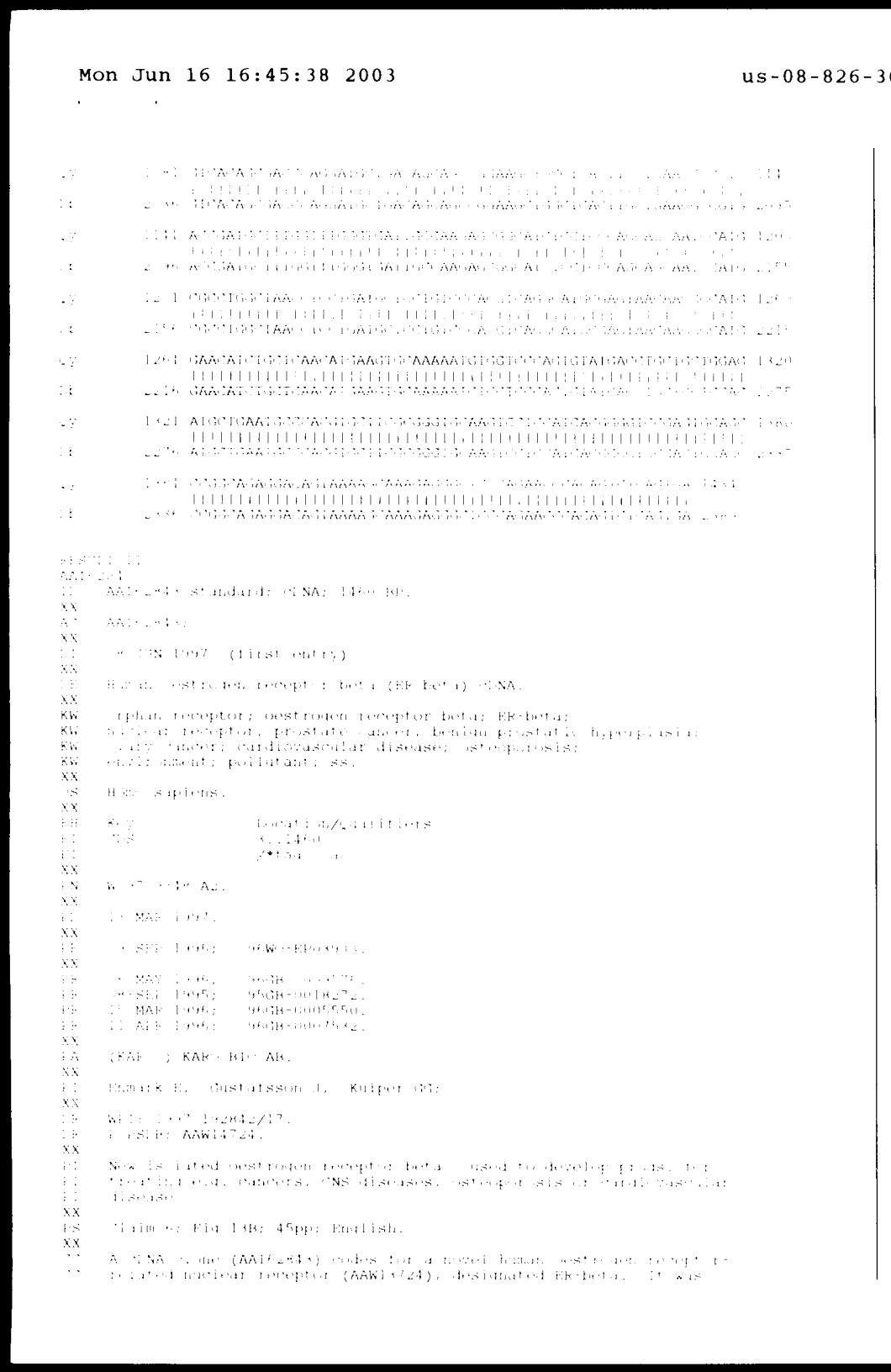
Qy 1141 GATTCTTCTTCTACTCCTTTTATTCCTATTACTACTGAAGGTGAAAATGGTTTTAAAAAT 1200
Db ... GATTCTTCTTCTACTCCTTTTATTCCTATTACTACTGAAGGTGAAAATGGTTTTAAAAAT 1787

Qy 1201 GAAGTTGAAGTTGATAATTCTCCTTTTTCTGGTTCTGTTTGTGTTGAAAAAGCTCAAGAA 1260
Db ... GAAGTTGAAGTTGATAATTCTCCTTTTTCTGGTTCTGTTTGTGTTGAAAAAGCTCAAGAA 1847

Qy 1261 GATTCTGGTTTTCCTGTTAATGATTATGGTACTCCTTGTCATGGTGCTACTTGTCCTCCT 1320
Db ... GATTCTGGTTTTCCTGTTAATGATTATGGTACTCCTTGTCATGGTGCTACTTGTCCTCCT 1907

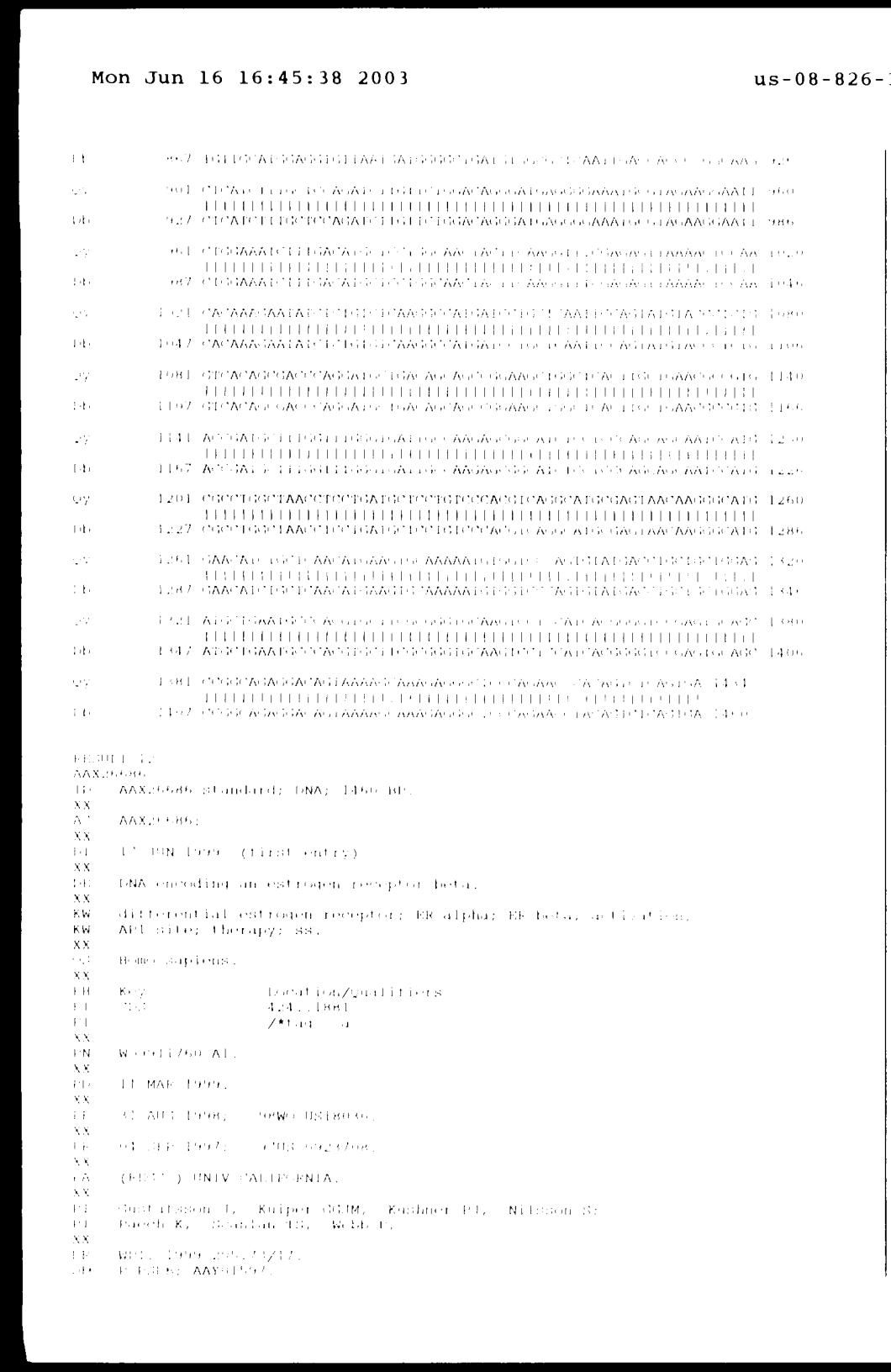
Qy 1321 GCTGTTGAAGGTCAAGCTGGTGCTGCTGTTCAAGCTGGTTTAGCTGCTGCTGCTGCTGCT 1380
Db ... GCTGTTGAAGGTCAAGCTGGTGCTGCTGTTCAAGCTGGTTTAGCTGCTGCTGCTGCT 1967

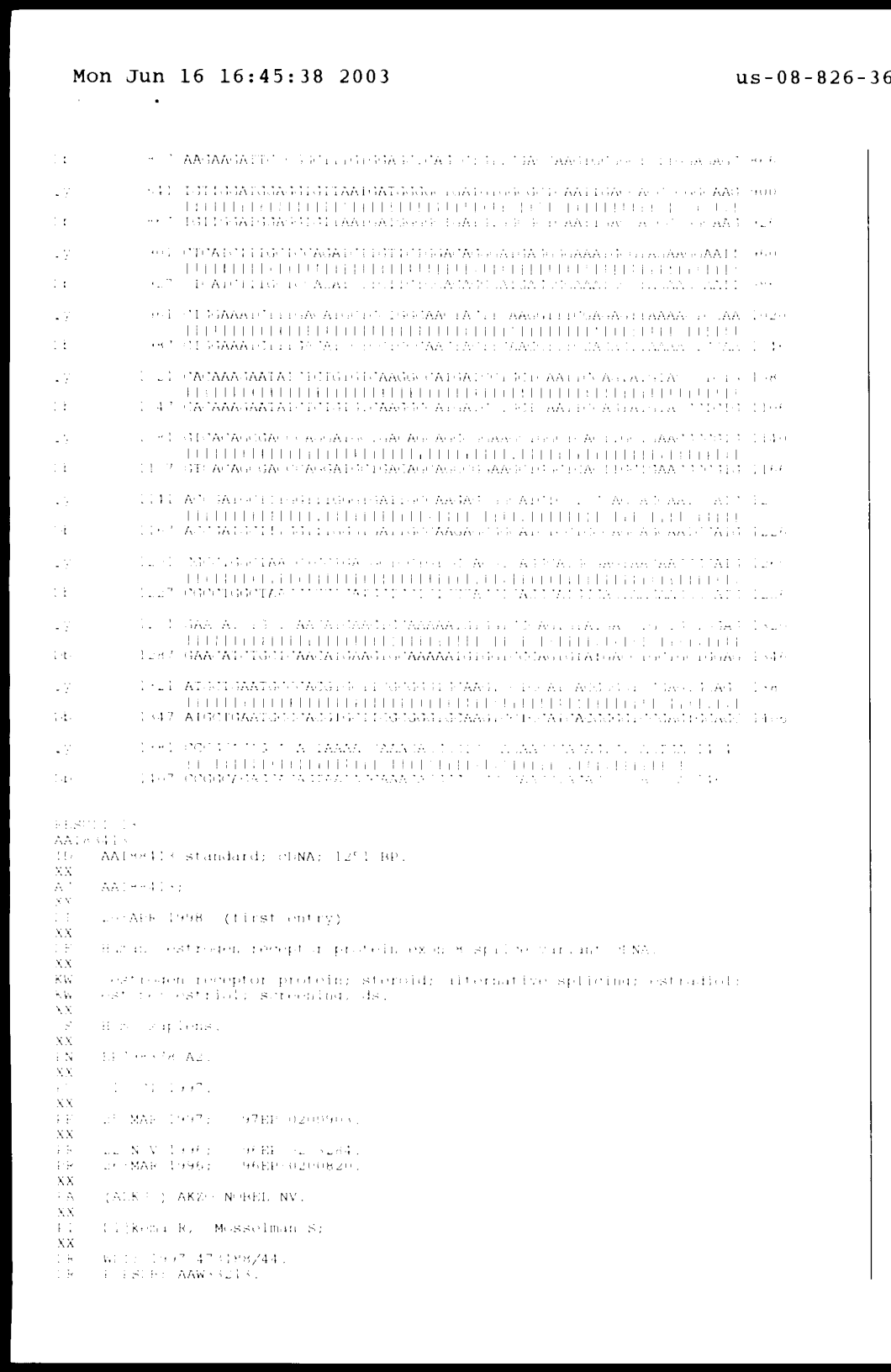




ID AAW14724 standard; cDNA; 1486 BP.
XX
AC AAW14724;
XX
DT 06-JUN-1997 (first entry)
XX
DE Rattus norvegicus oestrogen receptor beta (ER beta) cDNA.
XX
KW ligand receptor; oestrogen receptor beta; ER-beta;
KW nuclear receptor; prostatic cancer; benign prostatic hyperplasia;
KW fatty liver; cardiovascular disease; osteoporosis;
KW environmental pollutant; ss.
XX
OS Rattus norvegicus.
XX
RN [1]
RP 1-1486
RX MEDLINE; 96408722.
RX PUBMED; 8808628.
RA Kuiper G.G.J.M., Enmark E., Pelto-Huikko M., Nilsson S., Gustafsson J.-A.;
RT "Cloning of a novel receptor expressed in rat prostate and ovary.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:5925-5930(1996).
XX
RN [2]
RP 1-1486
RA Enmark E., Gustafsson J.-A., Kuiper G.G.;
RT ;
RL Submitted (17-JUN-1996) to the EMBL/GenBank/DDBJ databases.
RL Erasmus University, Department of Medical Biochemistry and Biophysics,
RL Medical School, P.O. Box 1738, 3000 DR Rotterdam, NETHERLANDS.
XX
CC Novel isolated oestrogen receptor beta - used to develop drugs for
CC treatment of cardiovascular diseases, genital tumours, CNS diseases,
CC osteoporosis.
XX
FH Key Location/Qualifiers
FT CDS 1..1486
FT /product="ER beta"
XX
SQ Sequence 1486 BP; 345 A; 417 C; 428 G; 296 T; 0 other;

Query Match 100.0%; Score 1486; DB 2; Length 1486;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

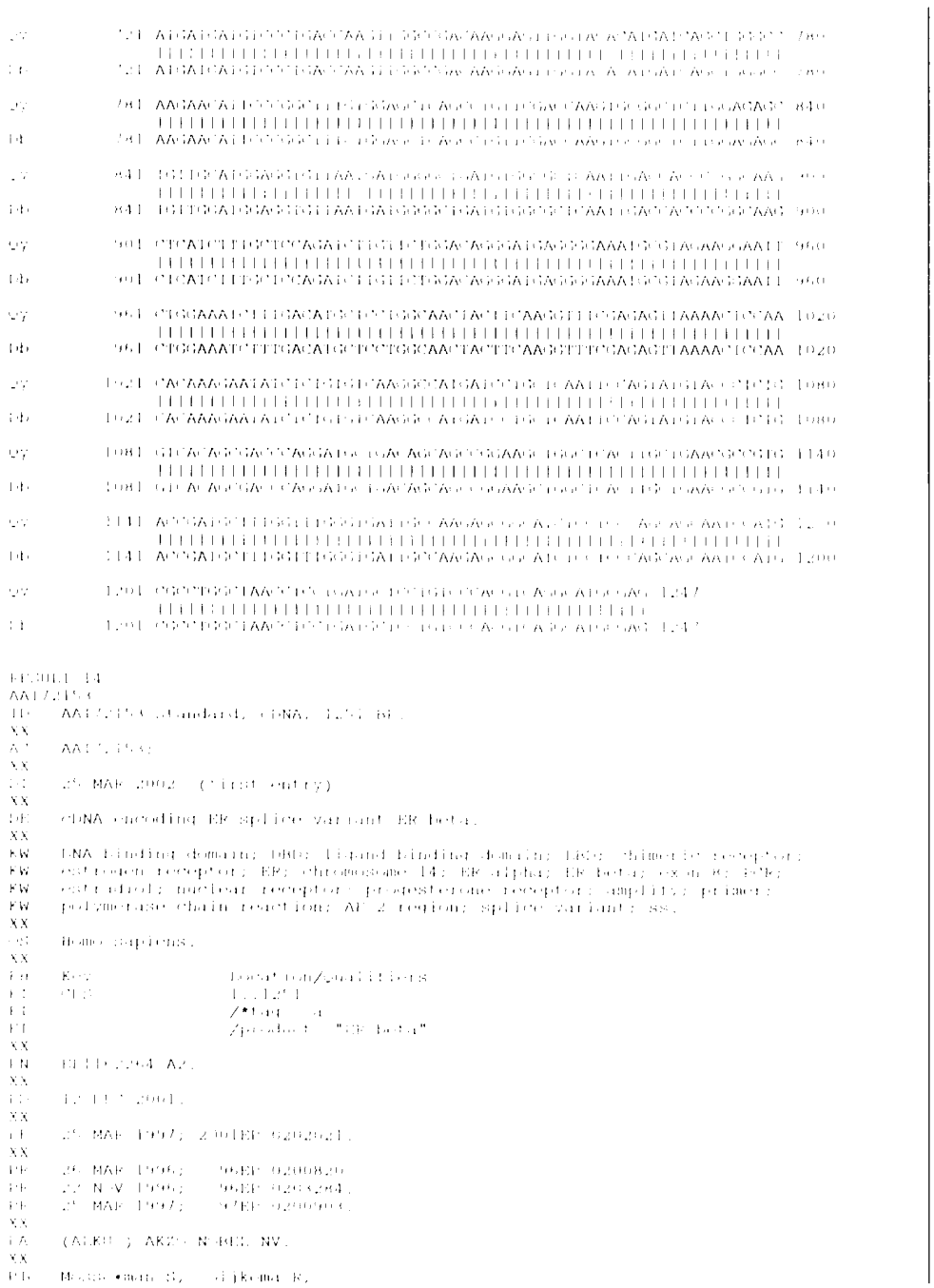


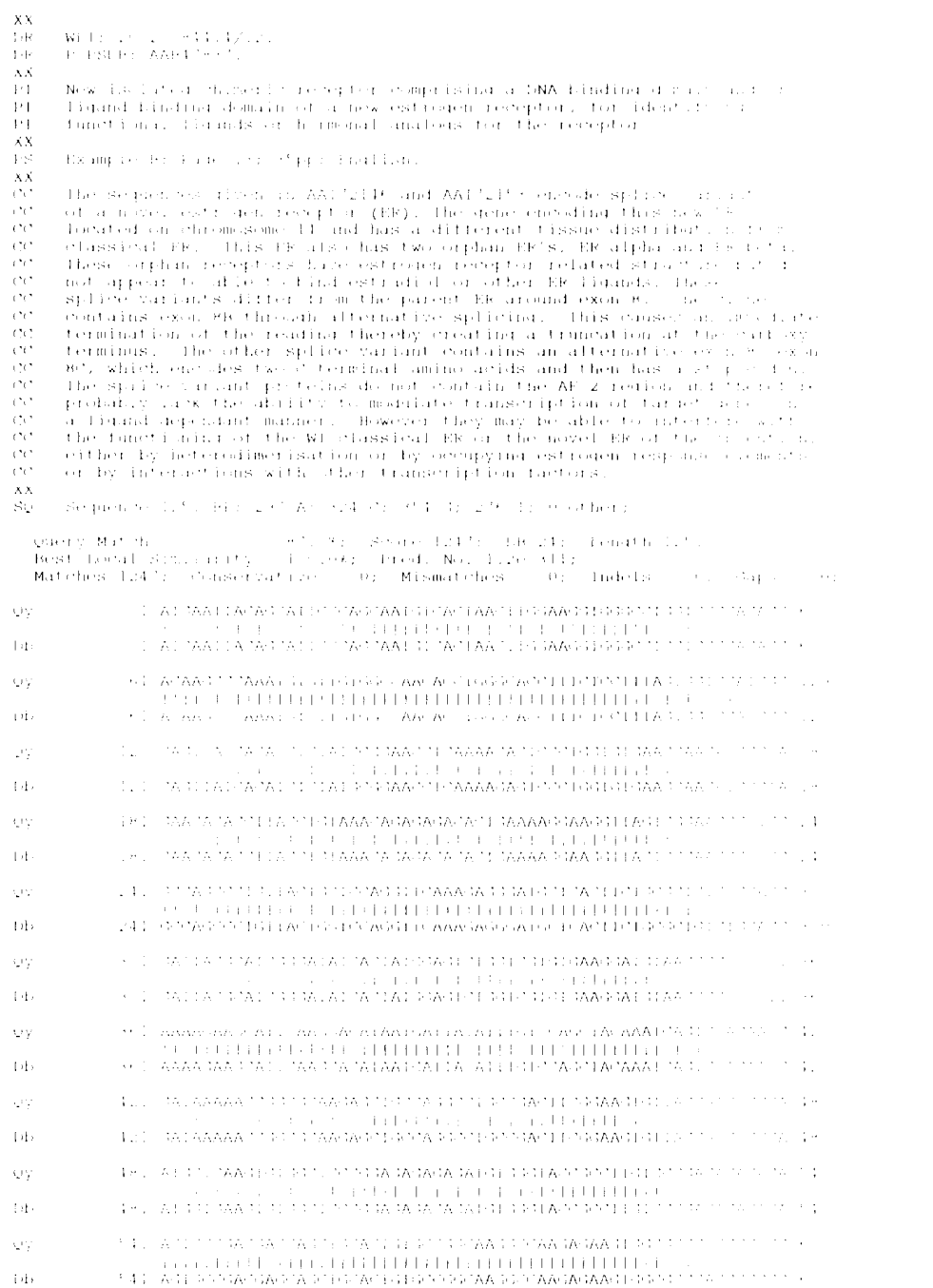


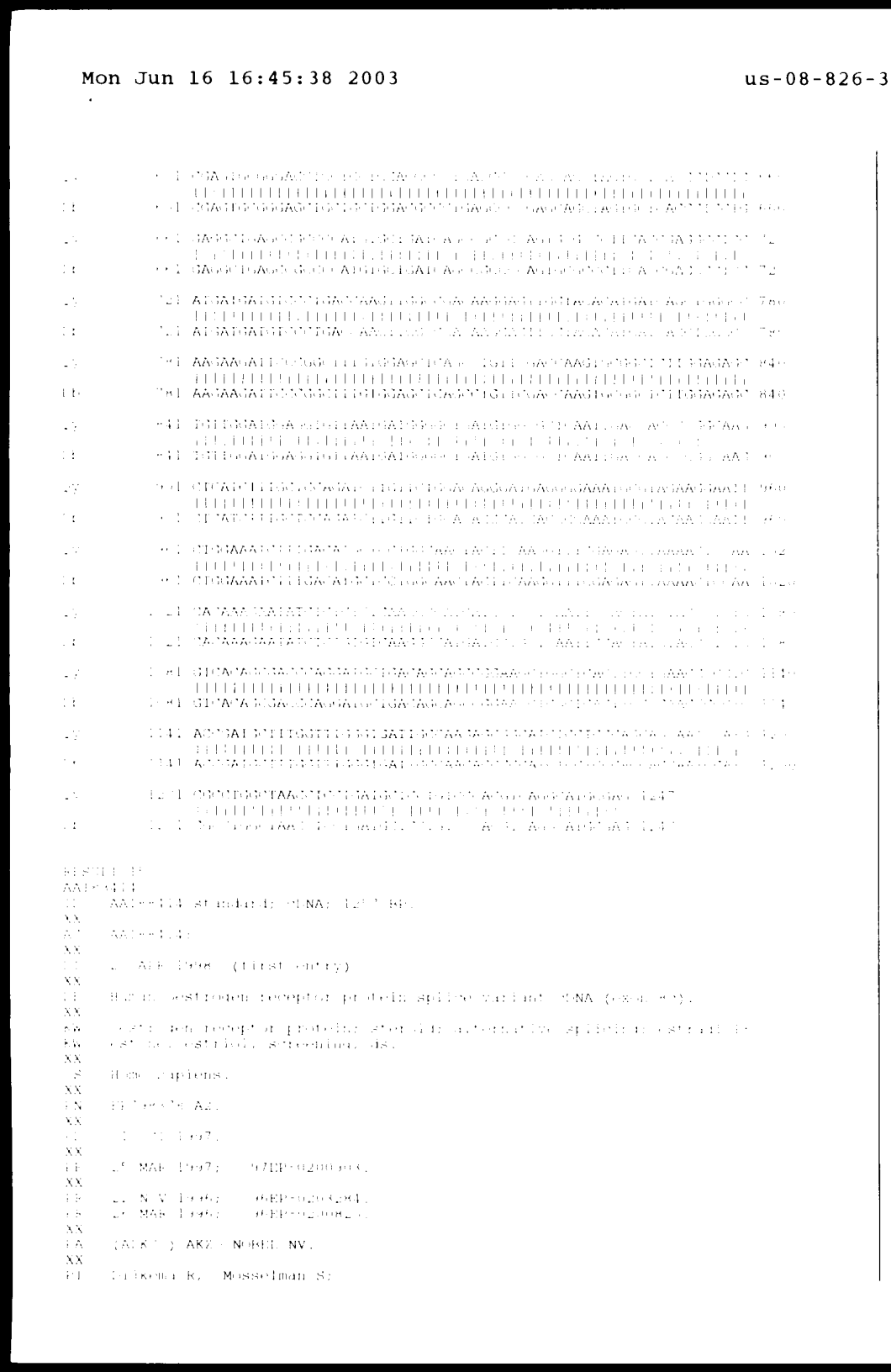
ID AAI99413 standard; cDNA; 1251 BP.
XX
AC AAI99413;
XX
DT 25-APR-1998 (first entry)
XX
DE Rat ret-estrogen receptor protein exon 4 splice variant cDNA.
XX
KW ret-estrogen receptor protein; steroid; alternative splicing; estradiol;
KW ret-estrogen receptor; screening; ds.
XX
OS Rattus sp.
XX
RN [1]
XX
RP 1-1251
XX
RX DOI 10.1006/MAMM.1996.0044; 97EF02008Mon.
XX
RA Kos N.V.;
RX PUBMED; 96EP-02008Mon.
XX
RA (AKZO) AKZO NOBEL NV.
XX
RT "Kos R, Mosselman S;
XX
RL Mol Cell Endocrinol 1996;117:47-59.
RL [ESR: AAW-521].

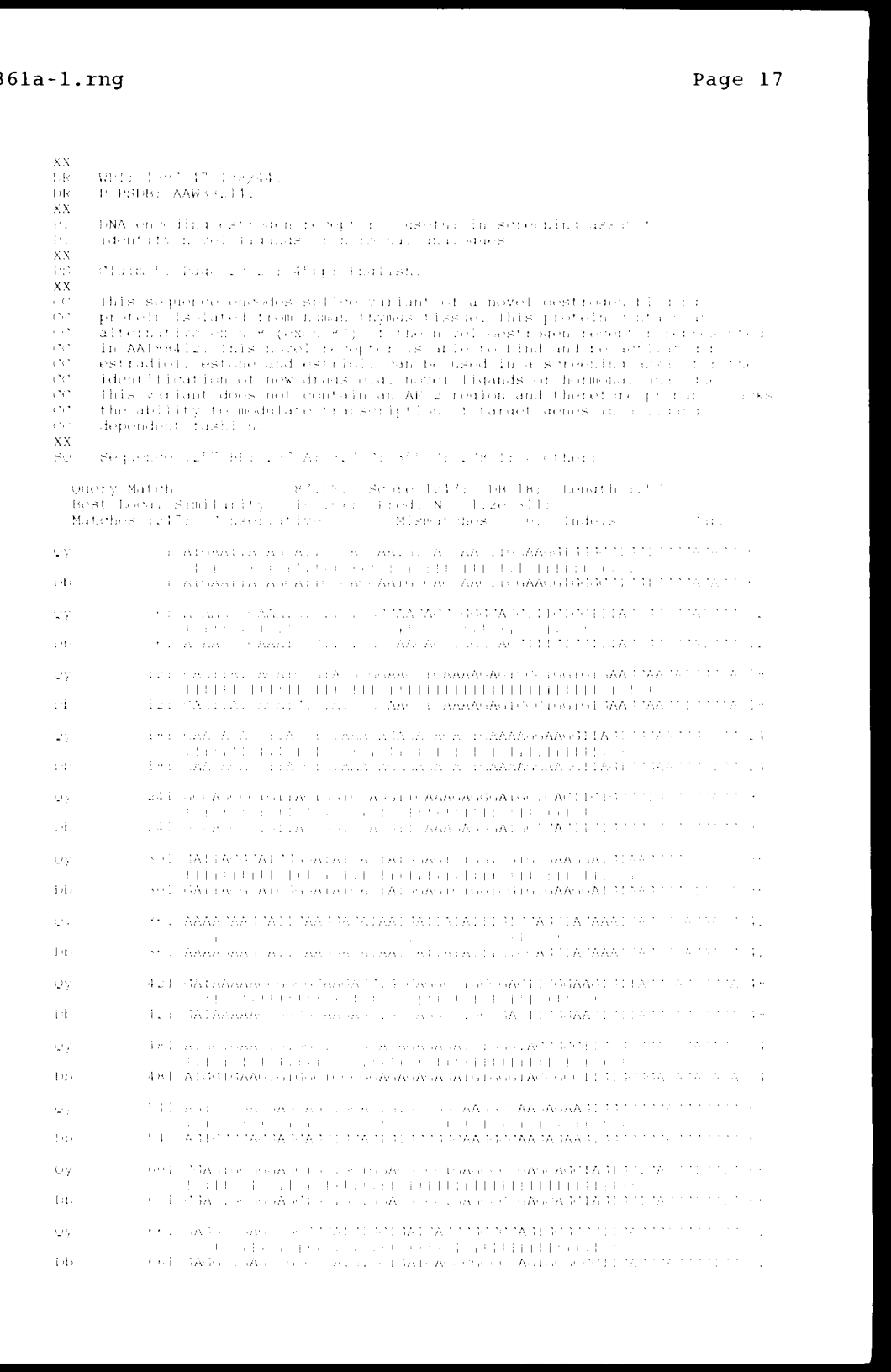
XX
DT DNA encoding estrogen receptor - useful in screening assays
DT identify novel ligands or hormonal analogues
XX
PS Claim 5; Page 21; 37pp; English.
XX
CC This partial sequence encodes a splice variant of a novel estrogen
CC binding protein detected by screening a human testis cDNA library.
CC The protein is expressed in testis, ovary, prostate (e.g. see AAI99415)
CC and testicular teratocarcinoma cells. The receptor is useful in
CC introduction of this exon-threonine alternative splicing event
CC the reading frame encoding the novel receptor is immediately
CC upstream of the stop codon. The receptor is useful in
CC identification of novel ligands or hormonal analogues
CC estrone and estriol, and can be used in a screening assay
CC identification of novel ligands or hormonal analogues
CC however this variant does not contain an AF-2 region and
CC prostate, it has the ability to bind transcription
CC ligands present in cells.
XX
SQ Sequence 1251 BP; 297 A; 414 C; 344 G; 270 T; 26 other;

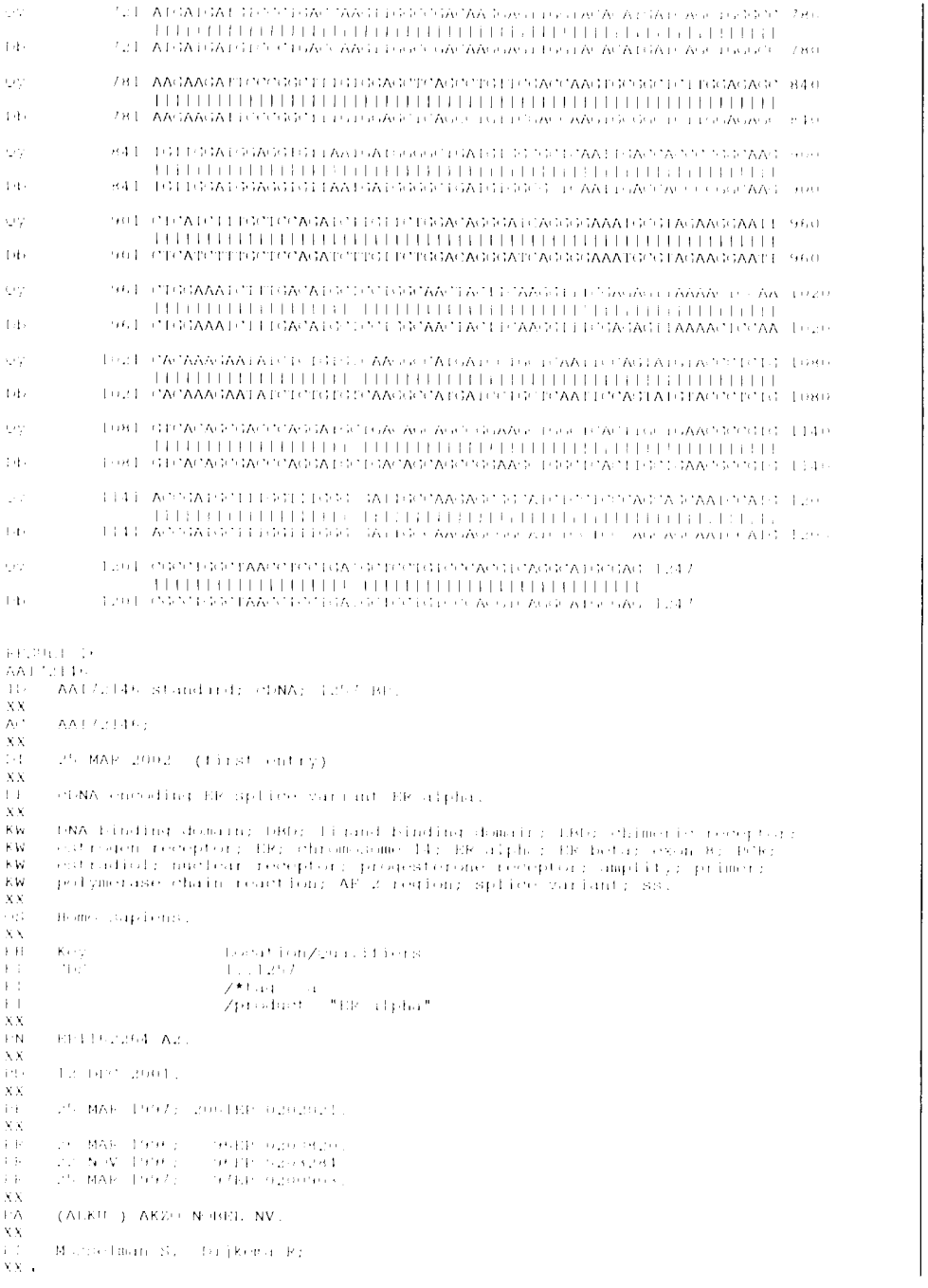
Query Match 23.6%; Score 1251; DB 19; Length 1251;
Best Local Similarity 100.0%; Pred. No. 1.2e-281;
Matches 1251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



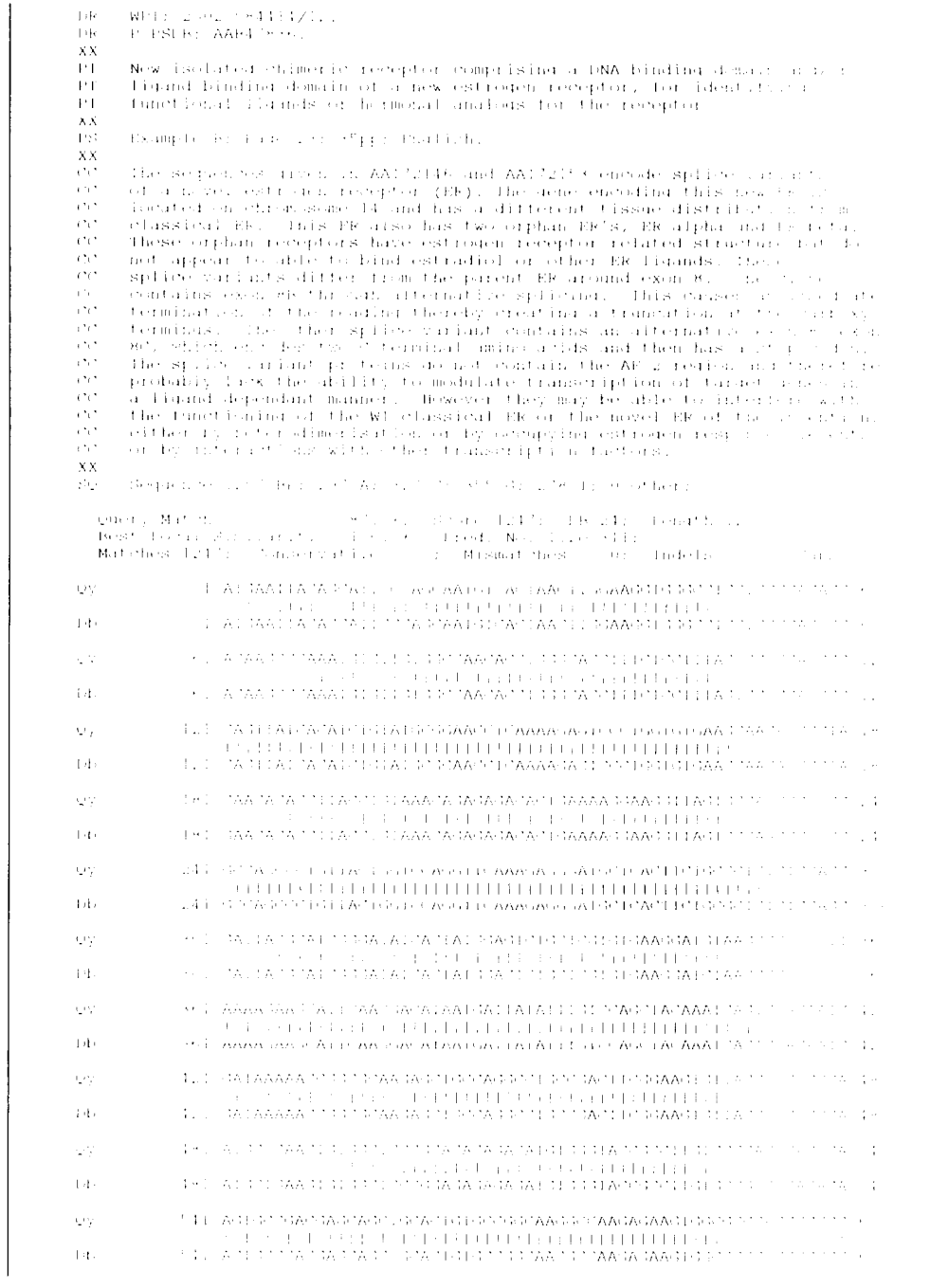






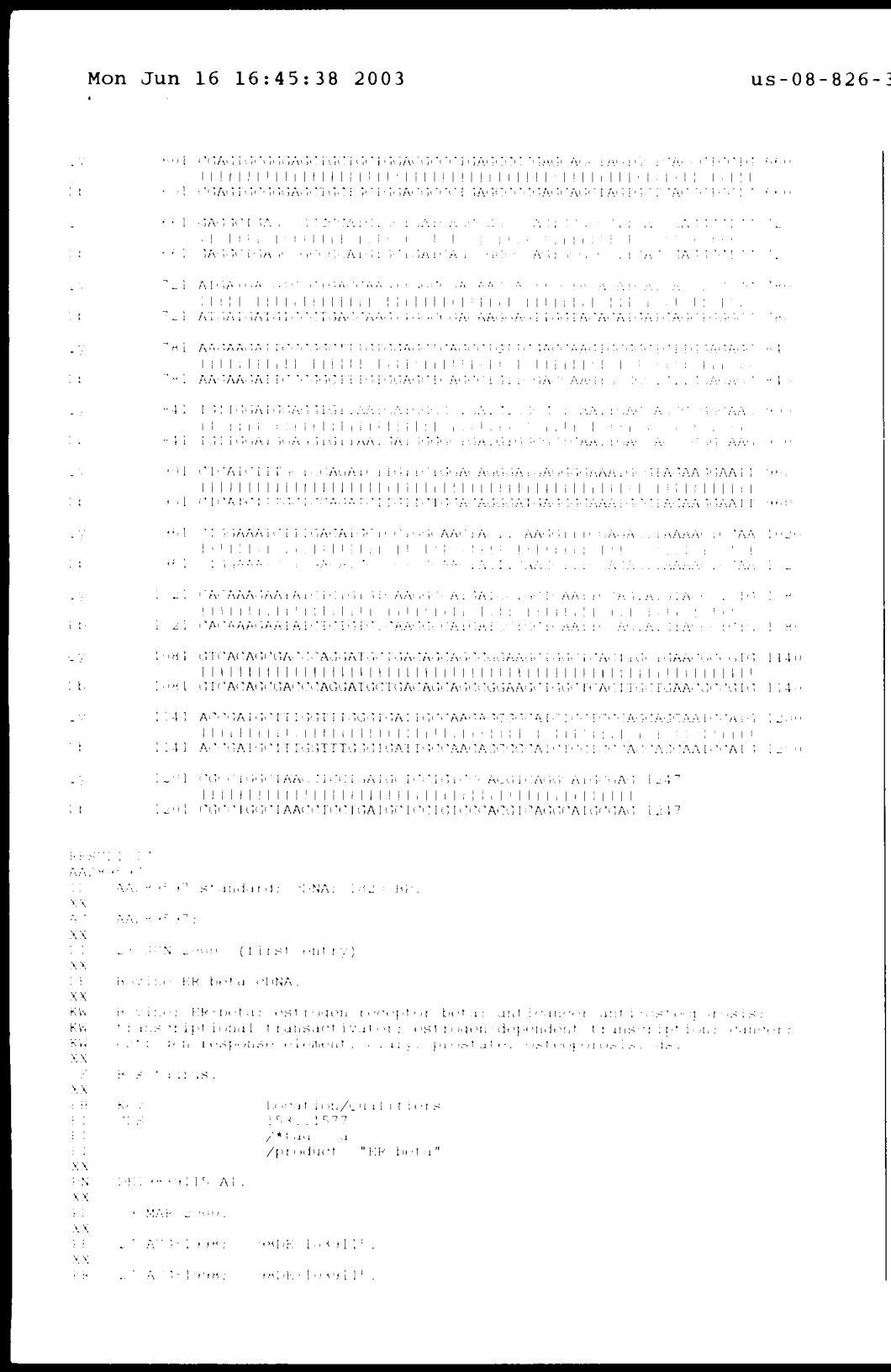


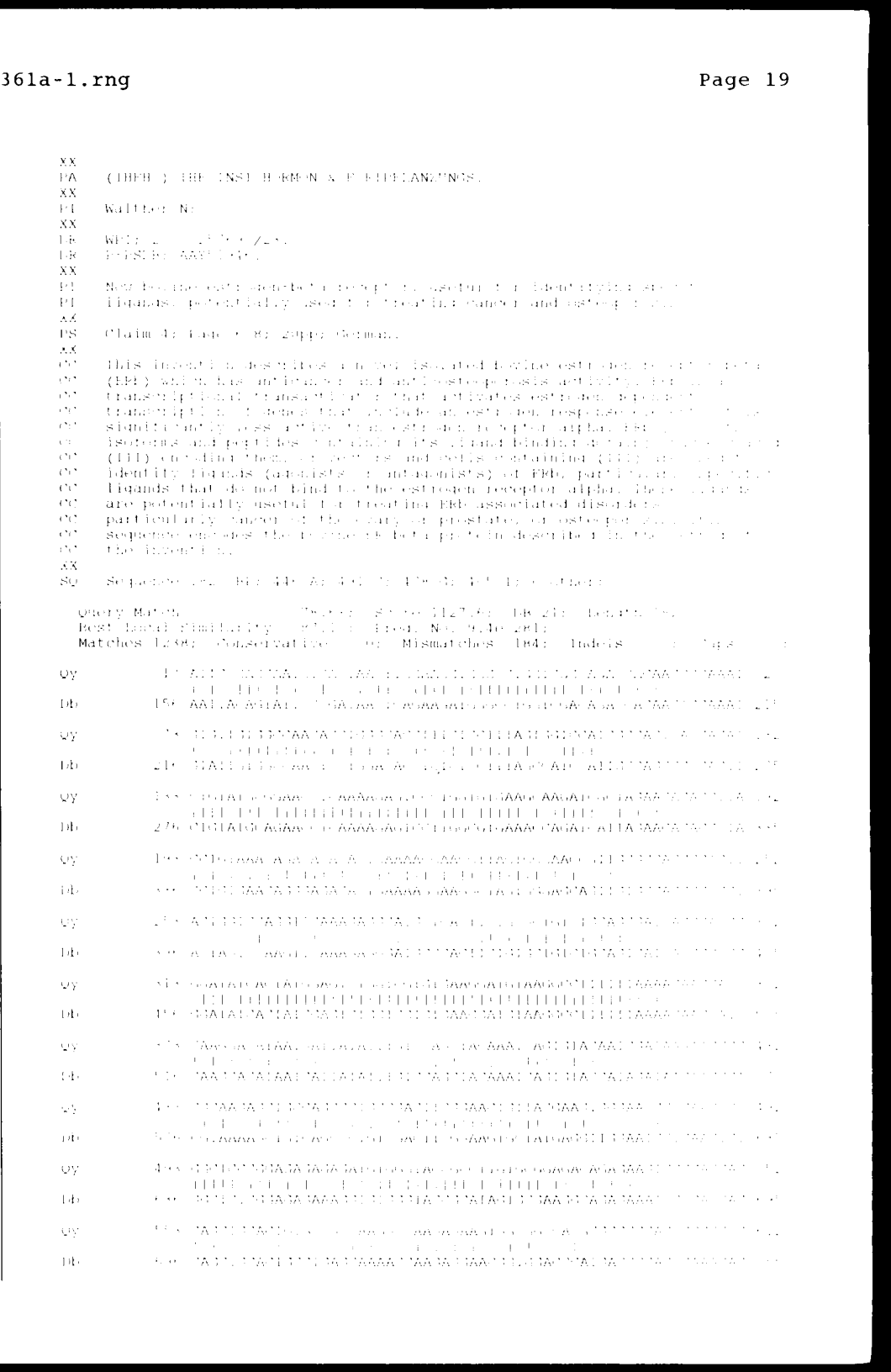
ID AAI72146 standard; cDNA; 1257 BP.
XX
AC AAI72146;
XX
DT 25-MAR-2002 (first entry)
XX
DE cDNA encoding ER splice variant ER alpha.
XX
KW DNA binding domain; DBD; ligand binding domain; LBD; chimeric receptor;
KW oestrogen receptor; ER; chromosome 14; ER alpha; ER beta; exon 8; hCBG;
KW oestradiol; nuclear receptor; progesterone receptor; amplify; primer;
KW polymerase chain reaction; AF-2 region; splice variant; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT "fo" 1..1257
FT /*tag = a
FT /product= "ER alpha"
XX
FN EP1102264-A2.
XX
PD 16-DEC-2001.
XX
PF 25-MAR-1997; 2001EP-0302021.
XX
PR 25-MAR-1996; 96EP-0200840.
PR 27-NOV-1996; 96EP-0203284.
PR 25-MAR-1997; 97EP-0200605.
XX
PA (ALEX-) AKZO NOBEL NV.
XX
PI Mittelman S, Pavlova E;
XX

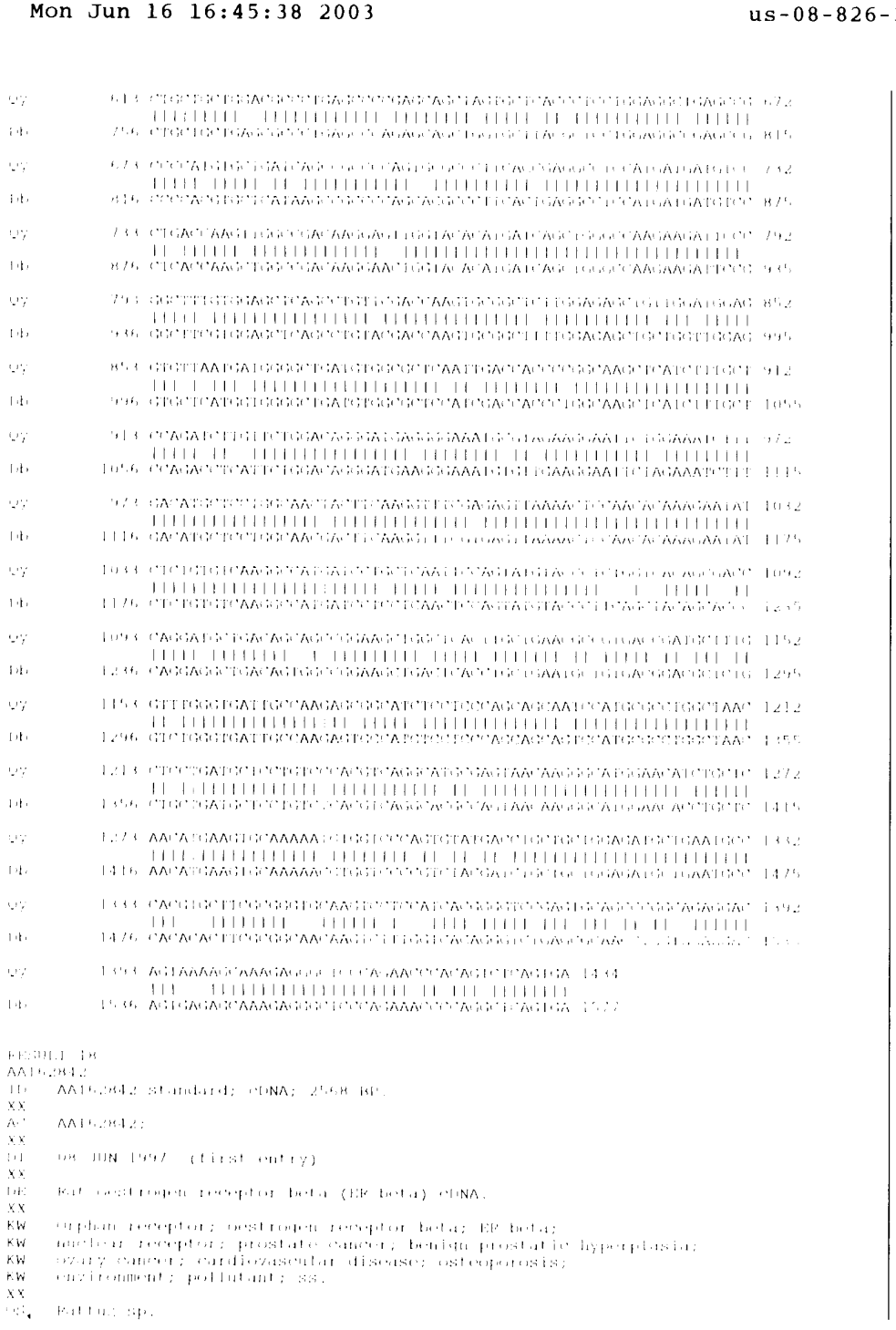


DR WPI; 2002-xxxxxx/11.
DR P-PSDB; AAB47xxx.
XX
PT New isolated chimeric receptor comprising a DNA binding domain or a
PT ligand binding domain of a new estrogen receptor, for identifying
PT functional ligands or hormonal analogs for the receptor
XX
PS Example 8; Fig 2; 47pp; English.
XX
CC The sequences from AA1-2146 and AA1-2258 encode splice variants
CC of a novel estrogen receptor (ER). The gene encoding this receptor is
CC located on chromosome 14 and has a different tissue distribution to a
CC classical ER. This ER also has two orphan ER's, ER alpha and beta.
CC These orphan receptors have estrogen receptor related structure but do
CC not appear to able to bind estradiol or other ER ligands. These
CC splice variants differ from the parent ER around exon 8. The variant
CC contains exon 8b (an alternative splicing). This causes a premature
CC termination of the protein thereby creating a truncated protein that
CC terminates. The other splice variant contains an alternative exon
CC 8c, which does not have a terminal amino acids and then has a short
CC The splice variant proteins do not contain the AF-2 region and so do
CC probably lack the ability to modulate transcription of target genes in
CC a ligand dependent manner. However they may be able to interfere with
CC the function of the WT classical ER or the novel ER of the invention,
CC either by ER dimerisation, or by competing estrogen response elements
CC or by interfering with other transcription factors.
XX
SQ Sequence 1257 BP; 317 A; 355 C; 327 G; 258 T; 0 other;

Query Match 100.0%; Score 1257; DB 21; Length 1257;
Best Local Similarity 100.0%; Pred. No. 0.0;
Matches 1257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

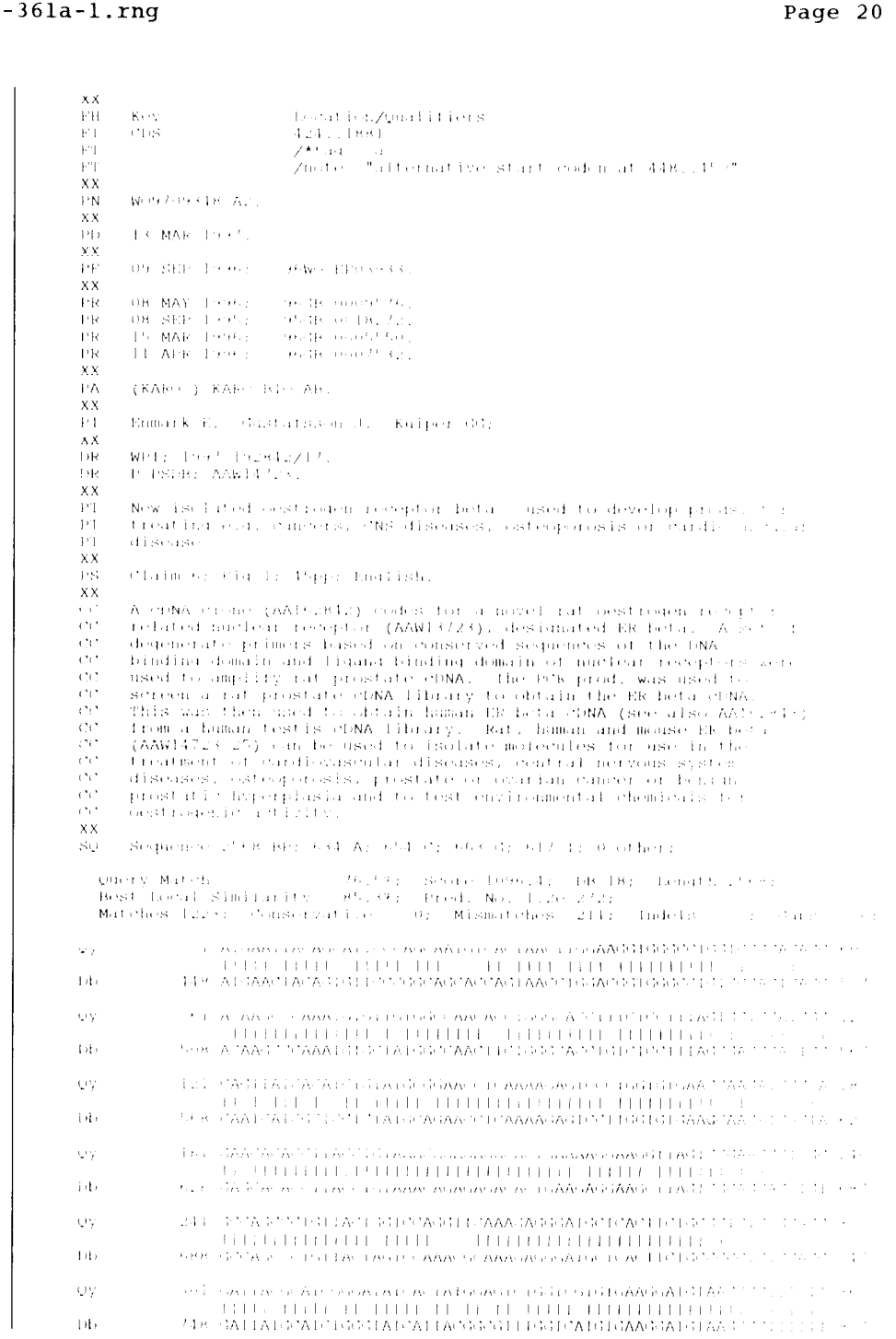


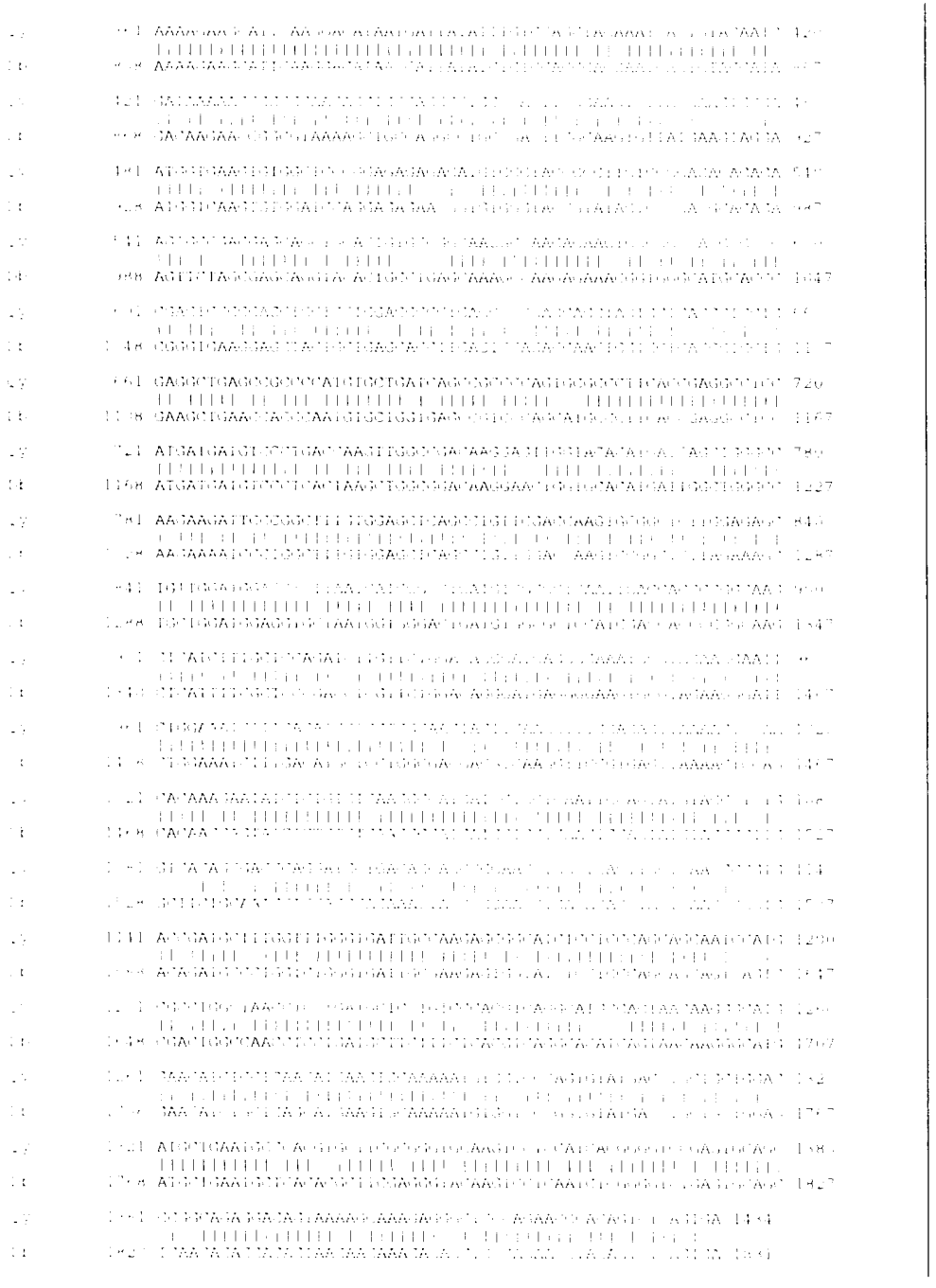




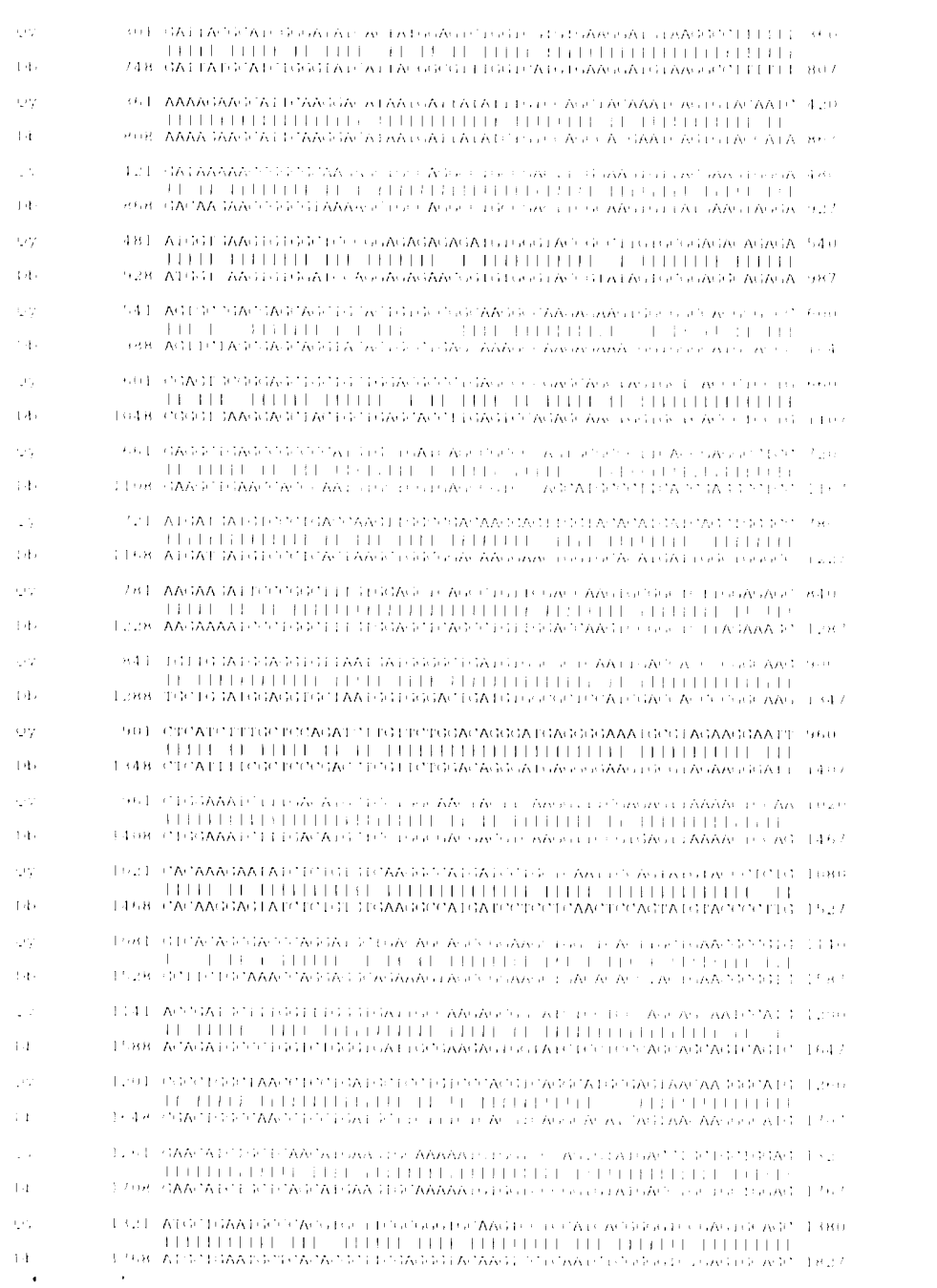
RESULT 18
AA162842
ID AA162842 standard; cDNA; 2568 BP.
XX
AC AA162842;
XX
DT 08-JUN-1997 (first entry)
XX
DE Rat oestrogen receptor beta (ER beta) cDNA.
XX
KW orphan receptor; oestrogen receptor beta; ER beta;
KW nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW ovary cancer; cardiovascular disease; osteoporosis;
KW environment; pollutant; ss.
XX
OS Rattus sp.

XX
FH Key Location/qualifiers
FT CDS 421..1891
FT /*tag= a
FT /note= "alternative start codon at 448..450"
XX
PN WO9729189-A2.
XX
PD 14-MAR-1997.
XX
PF 09-SEP-1996; 96WO-EP03933.
XX
PR 08-MAY-1996; 96GB-0009774.
PR 08-SEP-1995; 95GB-0018722.
PR 15-MAR-1996; 96GB-0005150.
PR 11-APR-1996; 96GB-0007432.
XX
PA (KARO-) KAROLINSKA INST.
XX
PI Enmark E; Gustafsson JA; Kuiper GG;
XX
DR WPI; 1997-164142/15.
DR P-PSDB; AAW14725.
XX
PT New isolated oestrogen receptor beta - used to develop prods. f
PT treating oest. cancers, CNS diseases, osteoporosis or cardi
PT disease
XX
PS Claim 9; Fig 1; 76pp; English.
XX
CC A cDNA clone (AA162842) codes for a novel rat oestrogen recept
CC related nuclear receptor (AAW14725), designated ER beta. A
CC degenerate primers based on conserved sequences of the DNA
CC binding domain and ligand binding domain of nuclear receptors were
CC used to amplify rat prostate cDNA. The PCR prod. was used to
CC screen a rat prostate cDNA library to obtain the ER beta cDNA.
CC This was then used to obtain human ER beta cDNA (see also AA1
CC from a human testis cDNA library. Rat, human and mouse ER beta
CC (AAW14725 25) can be used to isolate molecules for use in the
CC treatment of cardiovascular diseases, central nervous syste
CC diseases, osteoporosis, prostate or ovarian cancer or benign
CC prostatic hyperplasia and to test environmental chemicals for
CC oestrogenic activity.
XX
SQ Sequence 2568 BP; 634 A; 654 C; 663 G; 617 T; 0 other;



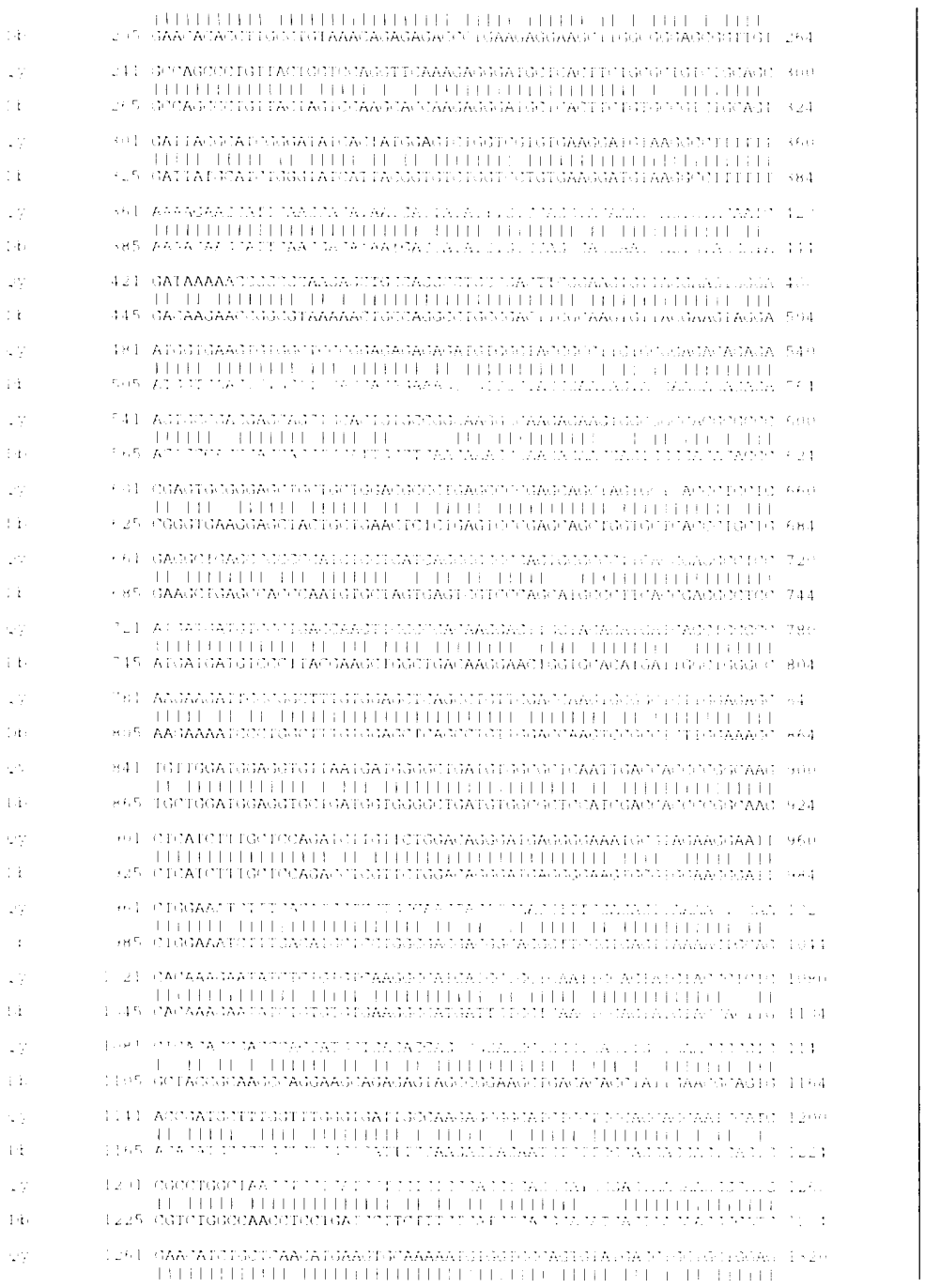


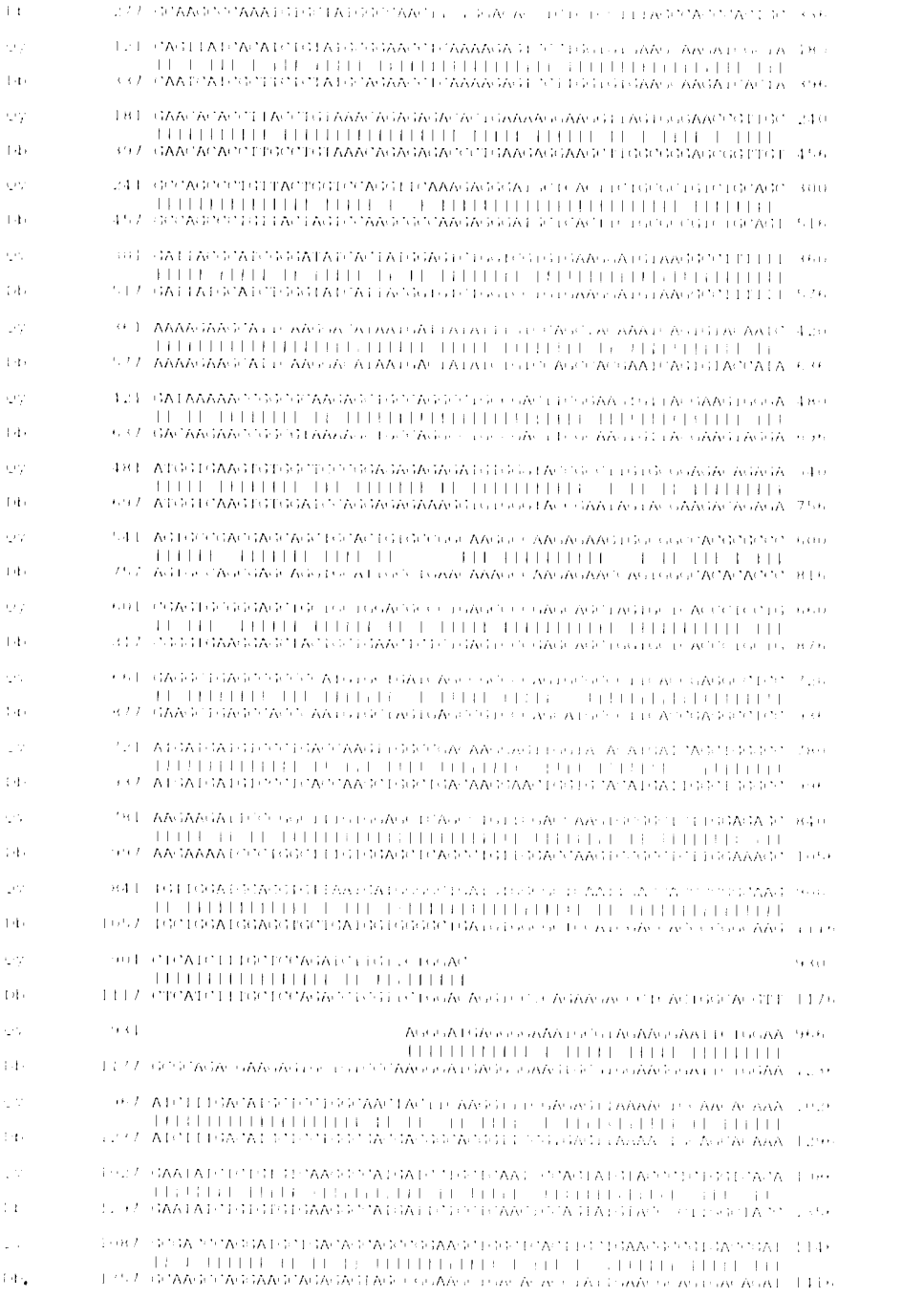
RESULT 19
AAX26685
ID AAX26685 standard; DNA; [illegible] BP.
XX
AC AAX26685;
XX
DT 17-JUN [illegible] (first entry)
XX
DE DNA encoding [illegible] mutation to optic beta.
XX
KW [illegible] optic beta; [illegible] mutation; [illegible] beta; [illegible].
KW Alcelaphinae.
XX
OS Kobus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..[illegible]
FT /**[illegible]
XX
PN WO9911765-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98WO-US018178.
XX
PR 04-SEP-1997; 97DE-1038178.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Gustafson G.J.; Kilpatrick D.M.; Kushner I.I.; Nilsson S.;
PI Powell K.; Spector D.S.; Wolfe H.;
XX
DR WPI; [illegible]-[illegible].
DR P-PSDB; AAY61500.
XX
PT Selective [illegible] test [illegible] with [illegible] [illegible] [illegible]



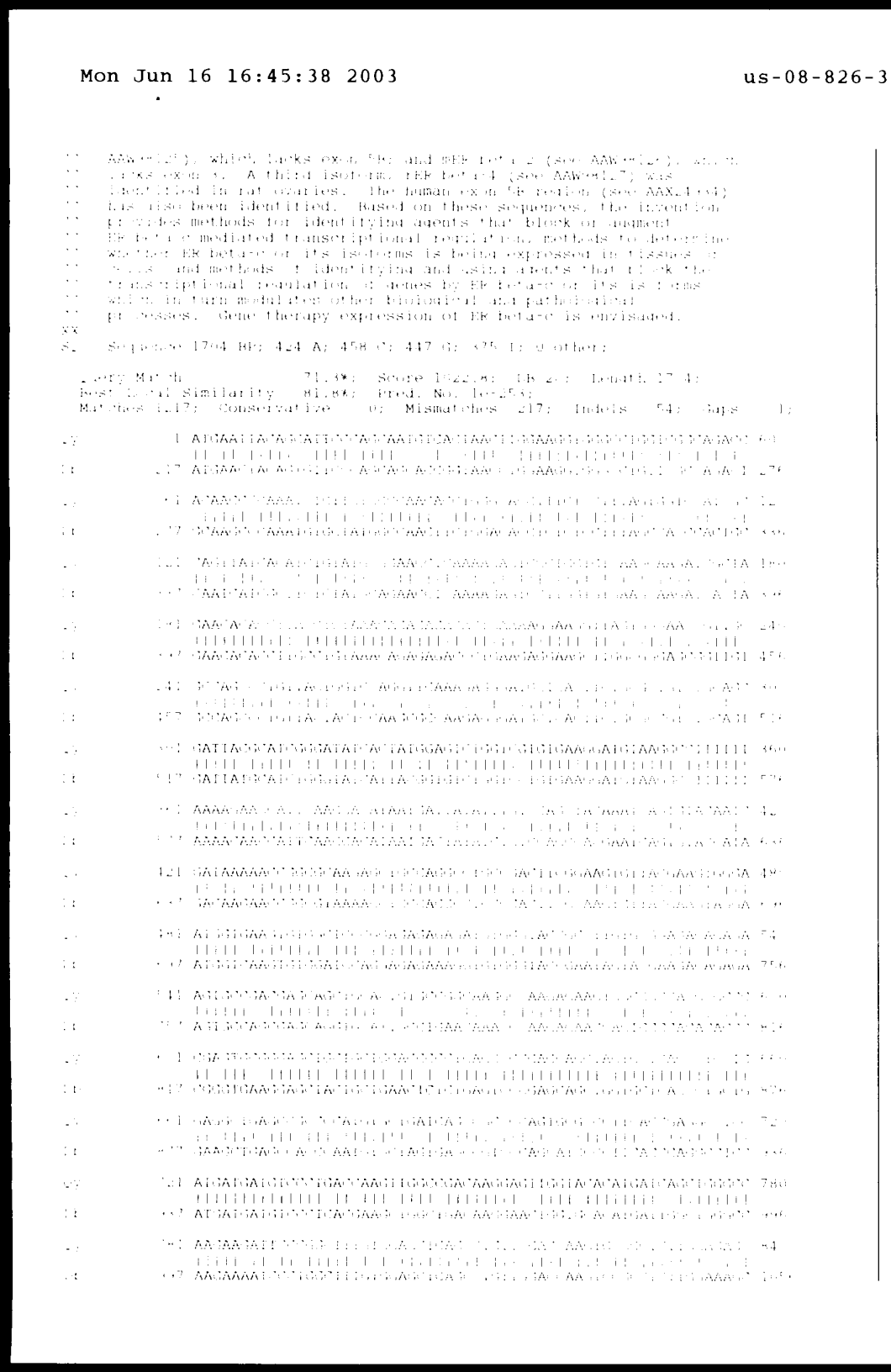
RESULT 20
AAI62841
ID AAI62841 standard; cDNA; 1755 BP.
XX
AC AAI62841;
XX
DT 08-JUN-2001 (first entry)
XX
DE Mouse oestrogen receptor beta (ER beta) cDNA.
XX
KW orphan receptor; oestrogen receptor beta; ER beta;
KW nuclear receptor; prostate cancer; benign prostatic hyperplasia;
KW ovarian cancer; CNS disease; osteoporosis;
KW cardiovascular disease.
XX
OS Mus sp.
XX
PN WO200164856-A2.
XX
PD 13-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US006406.
XX
PR 28-FEB-2000; 2000US-185553P.
PR 15-MAR-2000; 2000US-189450P.
PR 11-APR-2000; 2000US-196373P.
XX
PA (KARO-) KAROLINSKA INST.
XX
PI Gustafsson J-A, Warner M, Pettersson K, Delaunay F;
XX
DR WPI; 2001-574276/64.
DR P-PSDB; AAW61752.
XX
PT New isolated oestrogen receptor beta - used to develop products for
PT treating ovarian cancers, CNS diseases, osteoporosis or cardiovascular
PT disease.
XX
PS Claim 1; 114 pp; English.
XX
CC A cDNA (or its (AAI62840) protein or a novel mouse oestrogen
CC receptor beta (AAW61752), designated ER beta, has been
CC and mouse ER beta (AAW61752) can be used to isolate novel
CC use in the treatment of cardiovascular diseases, central nervous
CC system diseases, osteoporosis, prostate or ovarian cancer or benign
CC prostatic hyperplasia and to test environmental chemicals for
CC oestrogenic activity.
XX
SQ Sequence 1755 BP; 447 A; 455 C; 455 G; 398 T; 0 other;

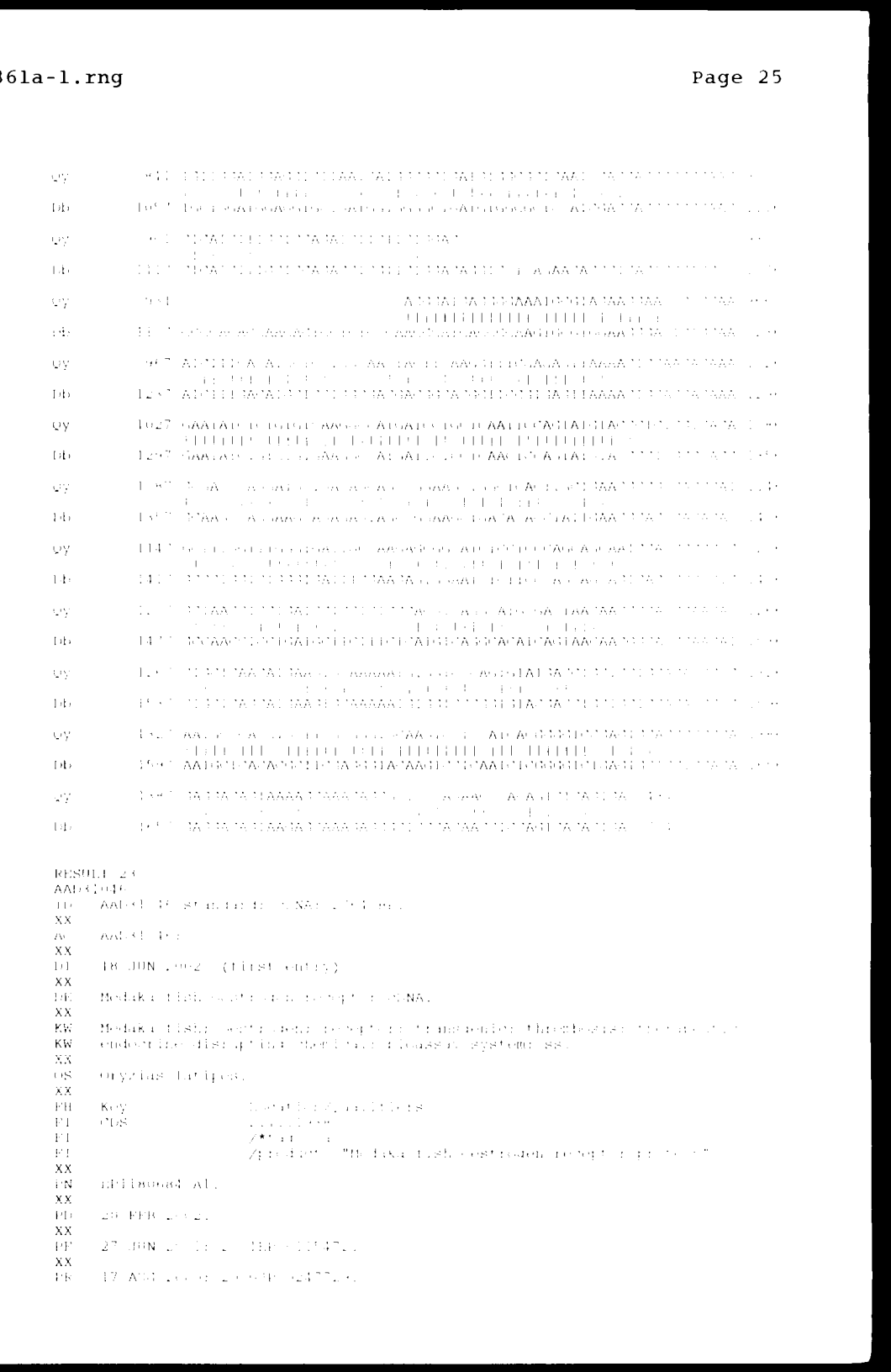
Query Match 100.0%; Score 1755; DB 20; Length 1755;
Best Local Similarity 100.0%; Pred. No. 0.0;
Matches 1755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

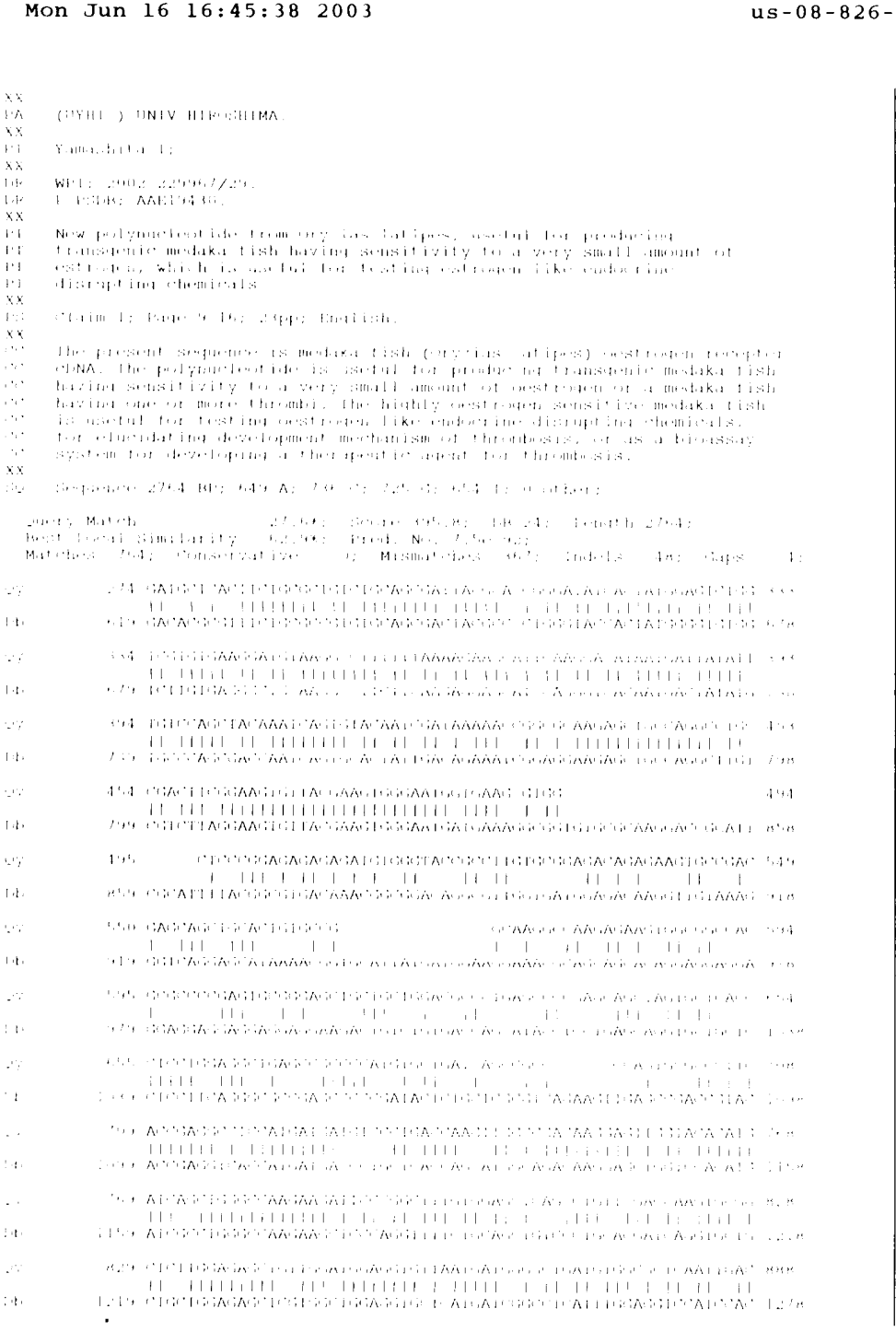


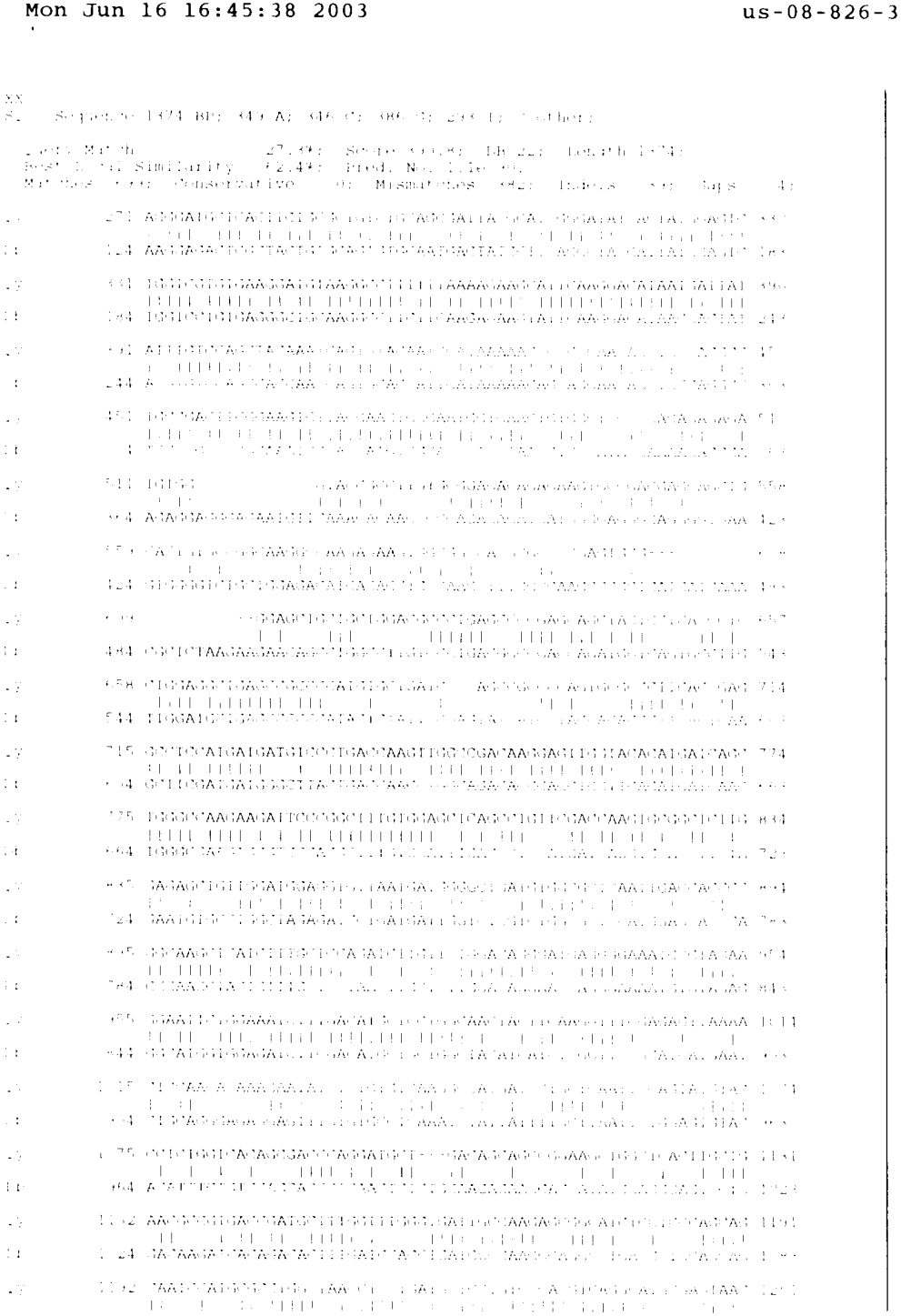


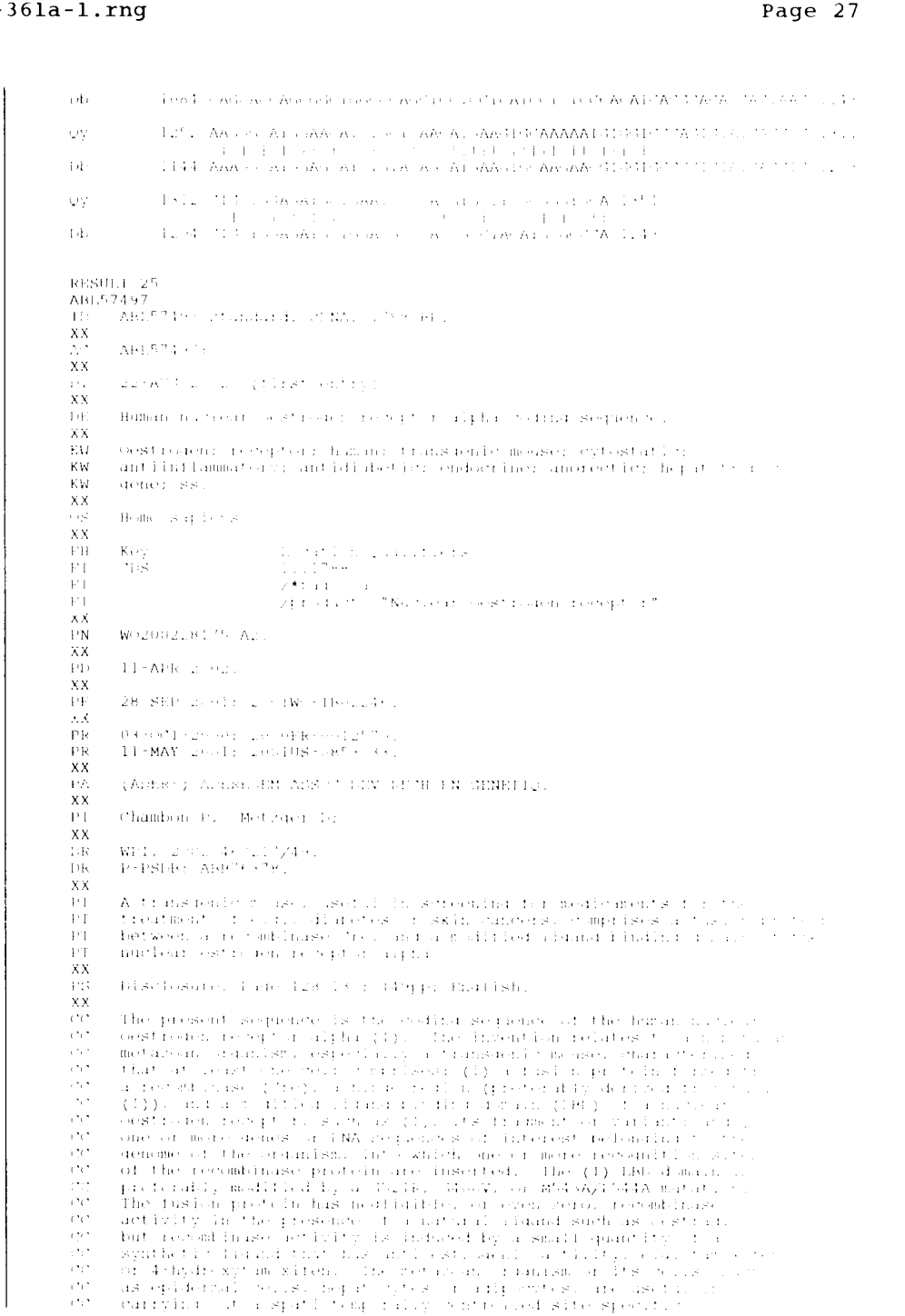
RESULT 22
AAX24***
ID AAX24*** standard; DNA; ** BP.
XX
AC AAX24***;
XX
DT ** JUN **** (first entry)
XX
DE Mouse complete coding sequence, receptor beta clone mERbeta 3.
XX
KW receptor; receptor; ER beta; ER beta 3; mouse; tumour; steroid;
KW therapy; ***.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT source ***..***
FT /*****="*"
FT /******="***"
XX
RN [*] *** ***
XX
RP ***-***
XX
RX DOI; ********.
RX PUBMED; ********.
XX
RA ********;
XX
RT "*****";
RL Submitted (**-***-****) ********.
XX
DR WPI; ******-**.
DR P-PSDB; AAW***.
XX
FT Human oestrogen receptor beta 3 - useful in treating and ***
FT tumours
XX
PS Example ** on page ***.
XX
CC The present sequence was ****** ** ***** and ** the routine ***
CC receptor beta 3 (ER beta 3) cDNA clone mER beta 3. ***
CC ******** ** ** ***** ** ** ** ** (see AAW***.). The ***
CC ******** ** ** ***** ** ** ** ** which ** ** ** ***
CC in ER beta 3 ** ** ** ** ** ** ** ** ** ** ** ** ***
CC mER beta 3 ** ** ** ** ** ** ** ** ** ** ** ** ***
CC ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ***
CC separate ** ** ** ** ** ** ** ** ** ** ** ** ** ***
CC mouse ** ** ** ** ** ** ** ** ** ** ** ** ** ***
CC spliced ** ** ** ** ** ** ** ** ** ** ** ** ** ***

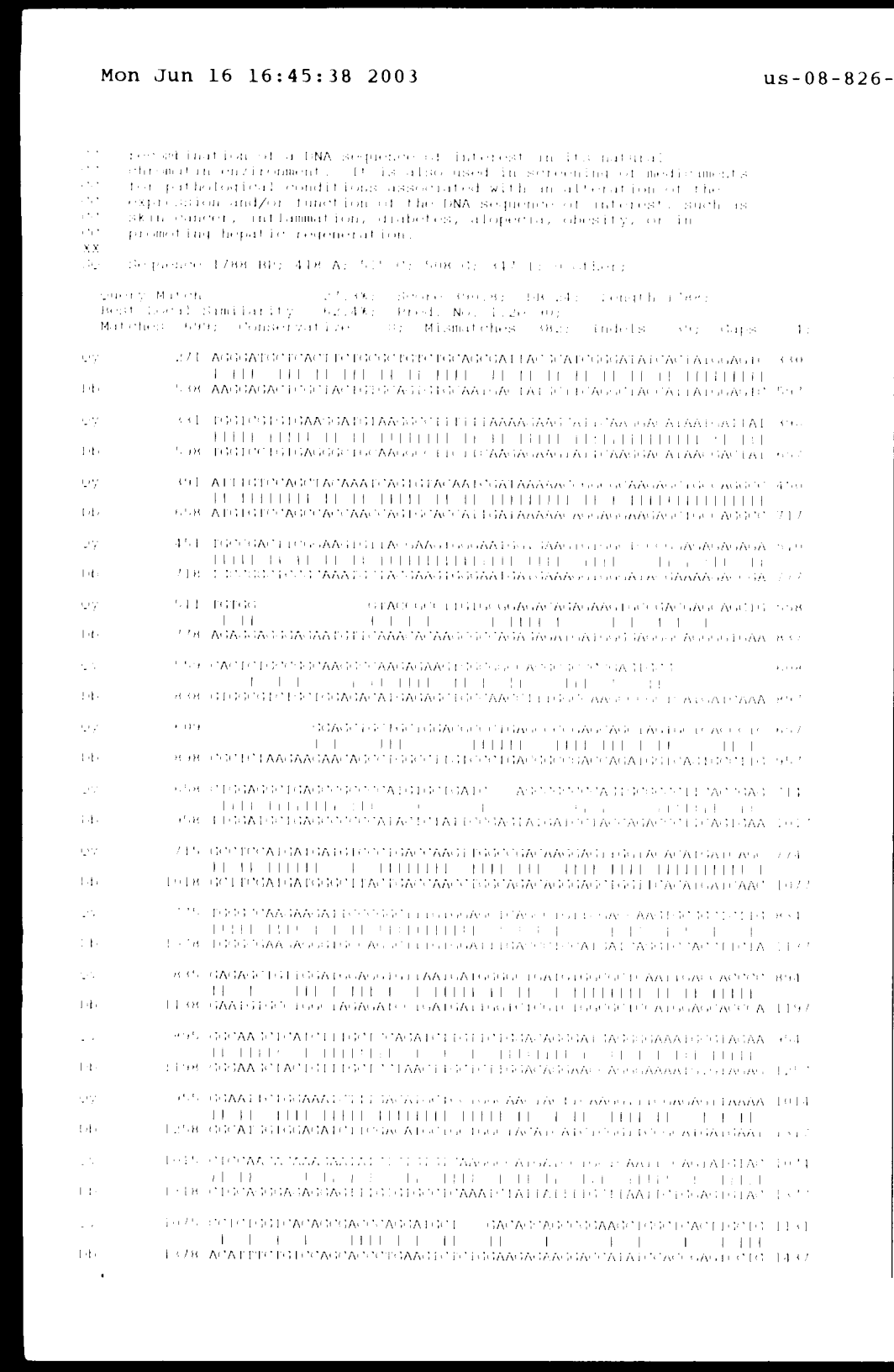


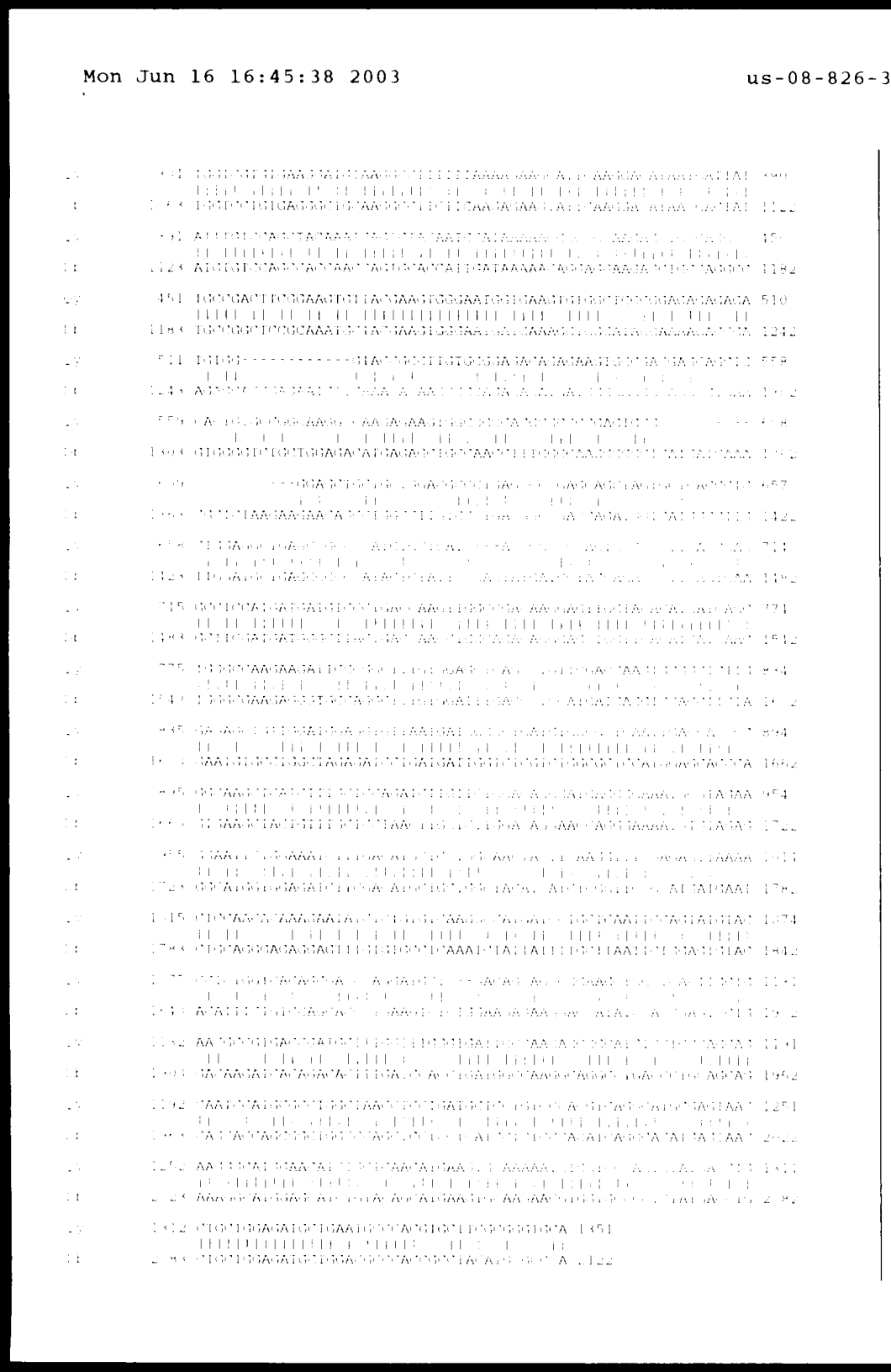




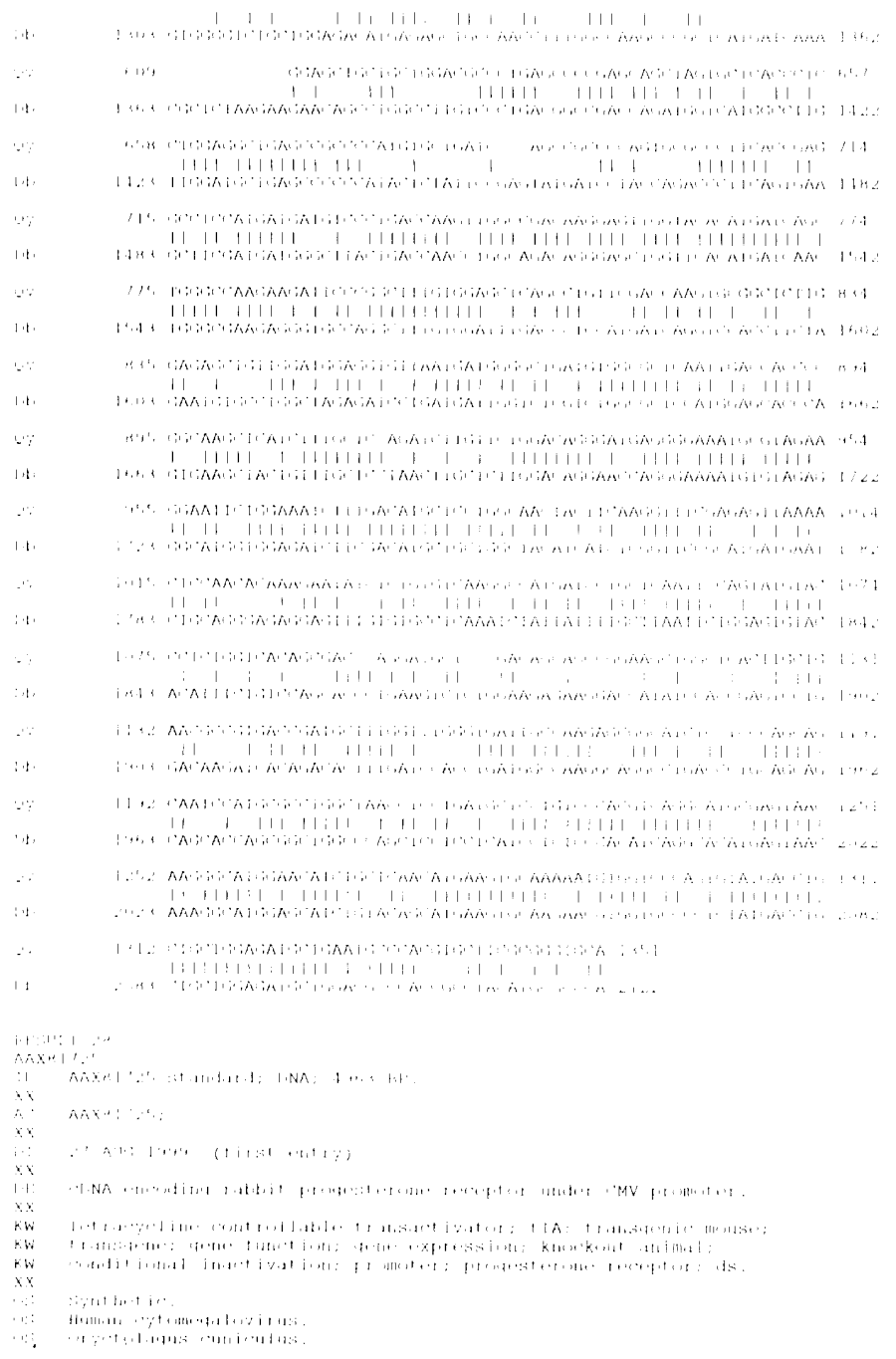


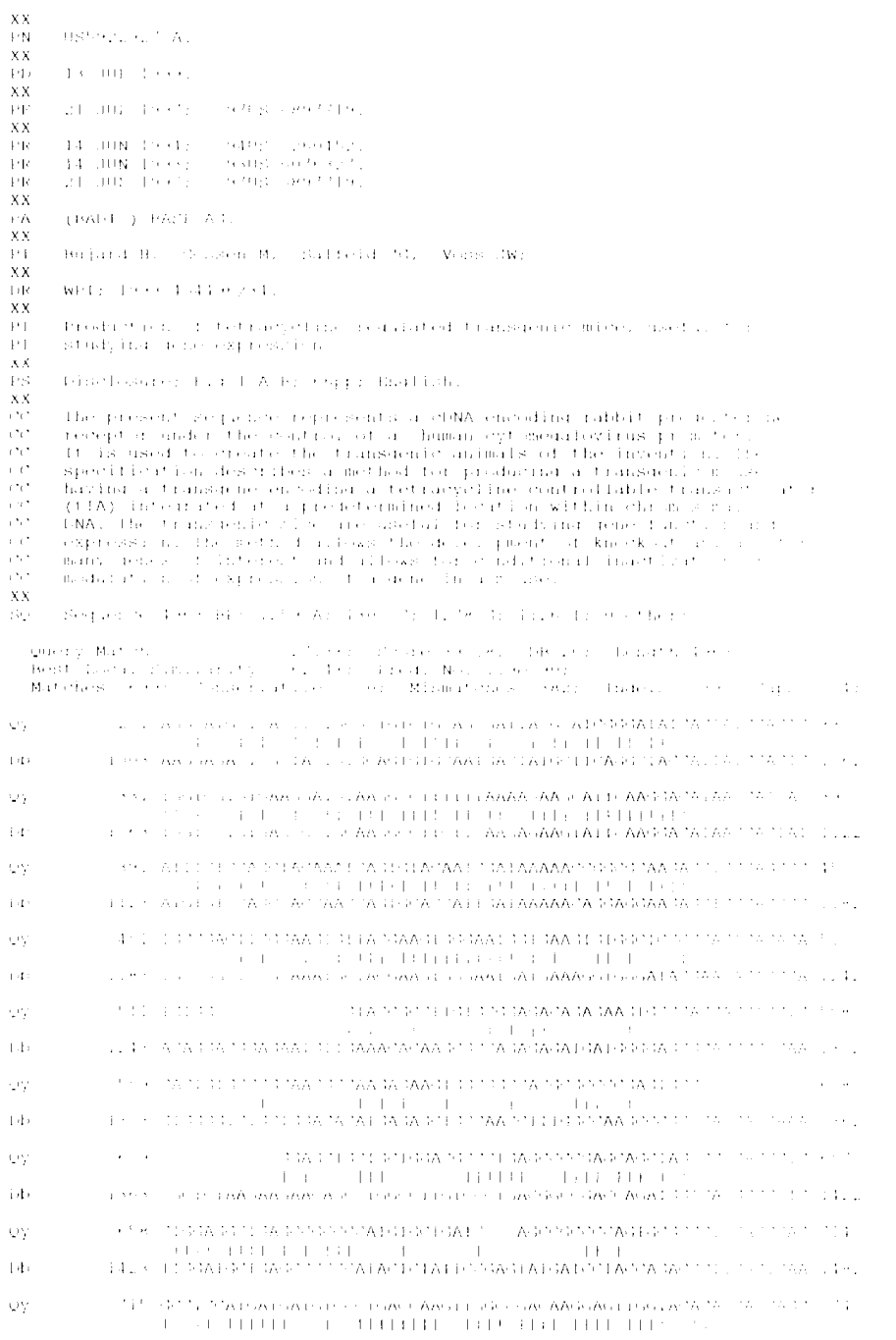


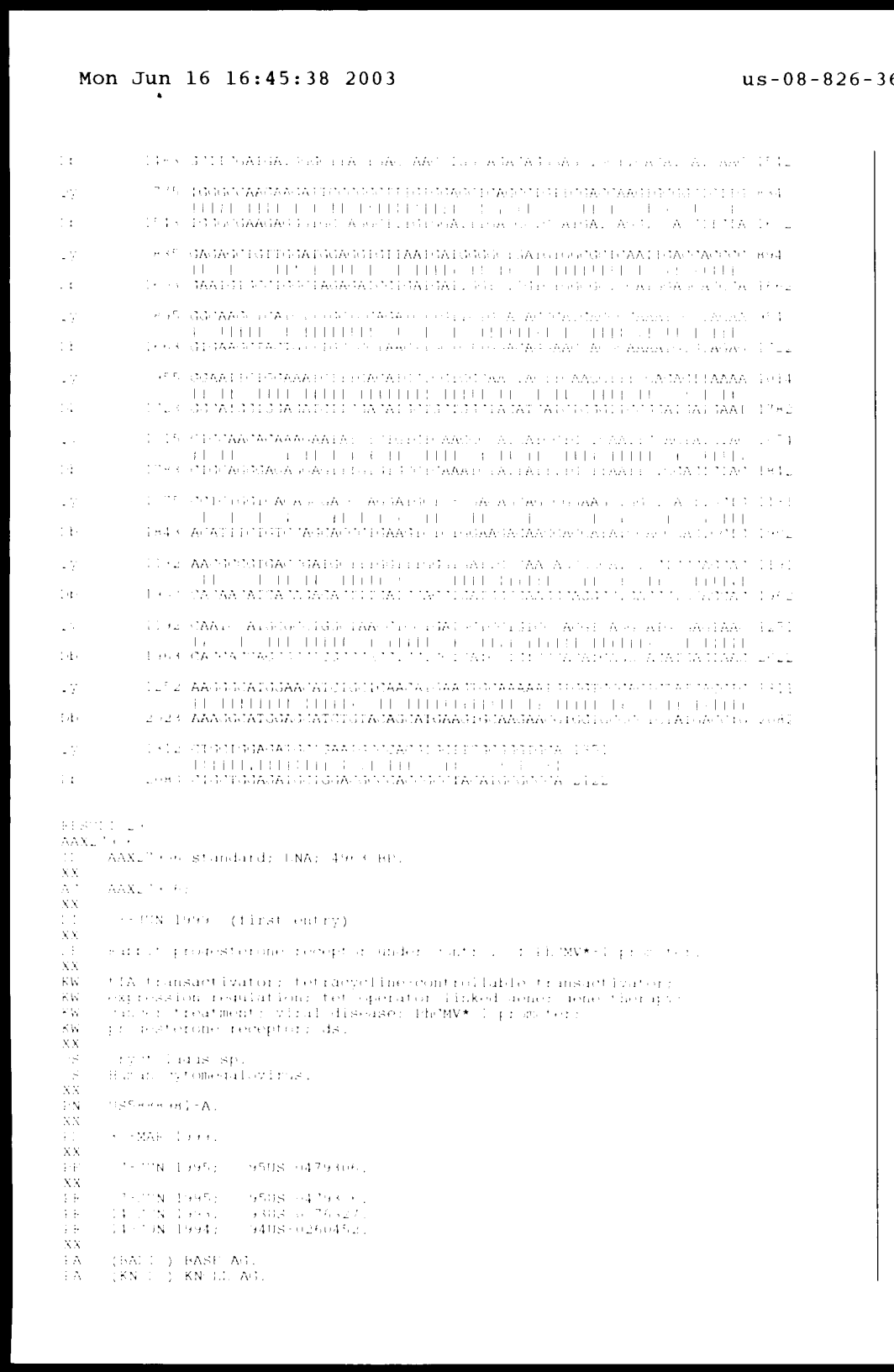




RESULT 27
AAQ76220
ID AAQ76220 standard; DNA; [illegible] BP.
XX
AC AAQ76220;
XX
DT 17-JUL-1995 (first entry)
XX
DE PhCMV*-1-controlled progesterone receptor.
XX
KW TIA; transactivator; tetracycline-controllable transactivator;
KW conditional inactivation; homologous recombination; gene expression;
KW gene regulation; gene therapy; tetracycline resistance; [illegible]
KW transgenic animal; PhCMV*-1; progesterone receptor; CMV;
KW promoter; receptor; ds.
XX
OS Human cytomegalovirus [unclassified].
XX
PN WO9429442-A.
XX
PD 22-DEC-1994.
XX
PF 11-JUN-1994; 94WO-US006611.
XX
PR 11-JUN-1993; 93US-0075727.
XX
PA (BAYE-) BAYER AG.
XX
PI Bujard H, Gossen M, Stratling WH, Voss JW;
XX
DR WPI; 1995-004172/01.
XX
PT Regulating eukaryotic transcription - by tetracycline-controllable transactivator
PT (TIA) - useful for conditional gene inactivation or modification
PT gene expression in transgenic cells and animals
XX
PS Disclosure; Page 34; [illegible] pp; English.
XX
CC The minimal promoter PhCMV*-1 is based on the human CMV
CC intermediate-early promoter (with enhancer region removed). Used
CC to express TIA transactivator in [illegible] cells. In combination with
CC [illegible] the promoter is used to
CC express TIA transactivator in [illegible] cells. Used to produce [illegible]
CC (also in AAQ76219). DNA is useful to [illegible]
CC PhCMV*-1.
XX
SQ Sequence 1454 BP; [illegible] A; [illegible] C; [illegible] G; [illegible] T; [illegible] other;





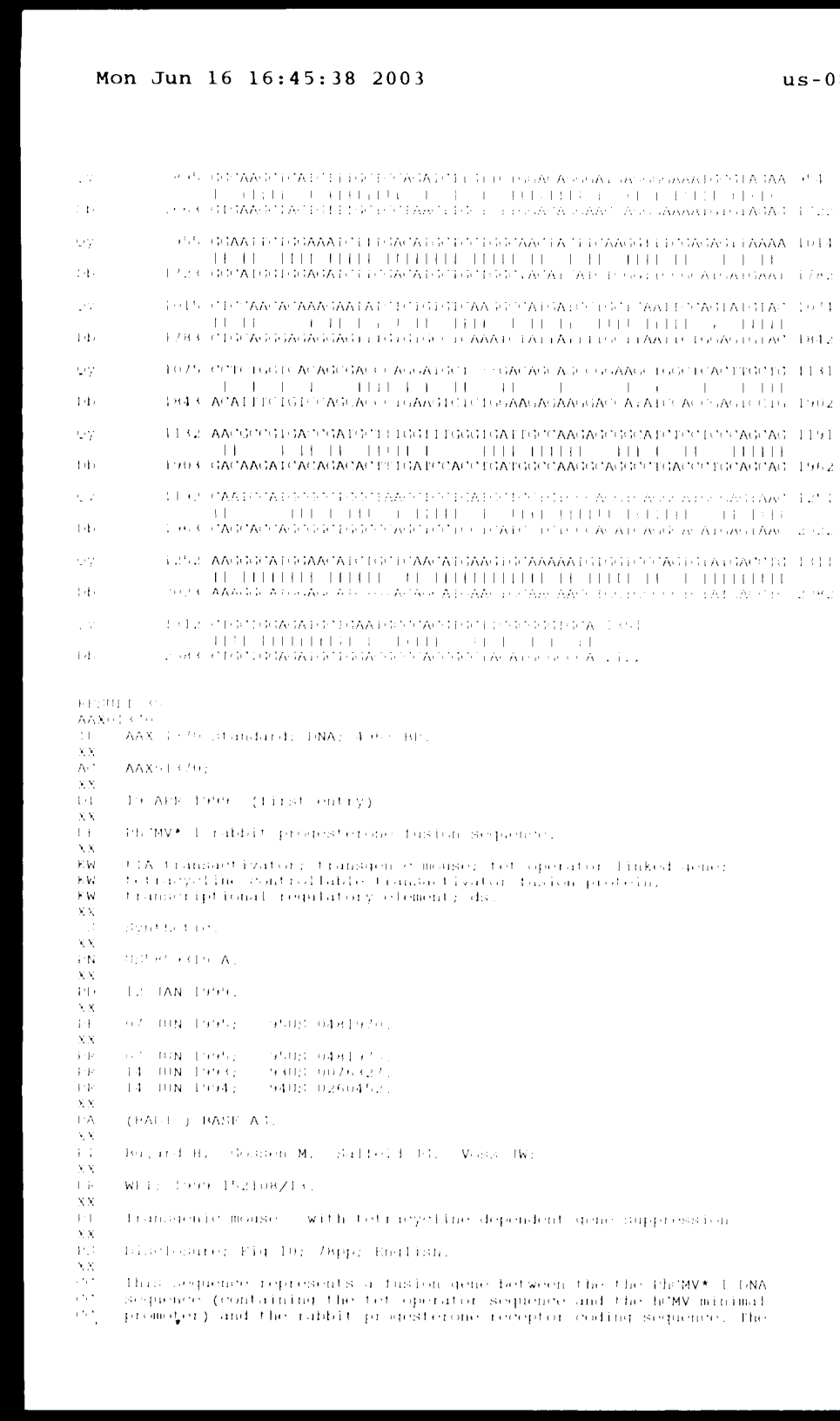


XX
PI Bejarat B Yasen M Shlomo ES Kizer We
XX
DR WPI; [illegible]
XX
PI New method for treating [illegible] expressing a tet operator binding [illegible]
PI useful in gene therapy applications for treatment of cancer
PI through overexpression of the tumor necrosis factor alpha gene
XX
PS Disclosure; Pg [illegible]; [illegible]
XX
CC This sequence represents the rabbit progesterone receptor gene
CC the entry is the MV* gene for separate the [illegible]
CC relates to a method of regulating expression of a target gene in a
CC gene therapy or a subject using a tetracycline control
CC transactivator (tTA) transactivator. The method can be used for
CC the level of gene expression and tetracycline has applications
CC therapy e.g. in cancer treatment (overexpression of a TNF alpha
CC INF alpha) or treatment of viral diseases by expressing [illegible]
CC negative viral transactivation proteins such as trans dominant
CC tet mutants of HIV. The system is also mediated to [illegible]
CC expressed [illegible] which has not been possible in prior art
CC systems.
XX
SQ Sequence [illegible] AA; [illegible] CA; [illegible] G; [illegible] T; [illegible] other;

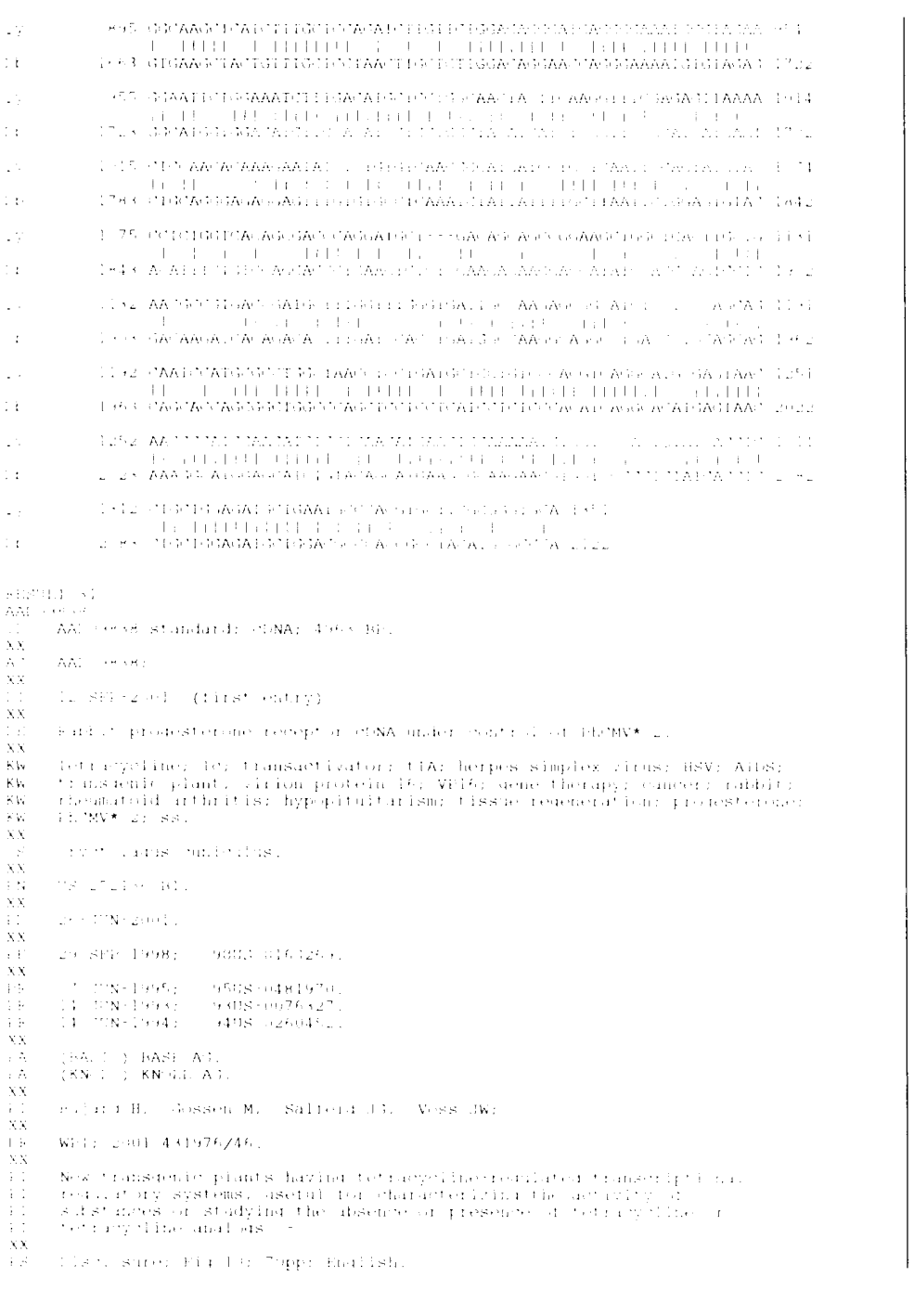
Query Match [illegible]; Score [illegible]; DB [illegible]; Length [illegible];
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

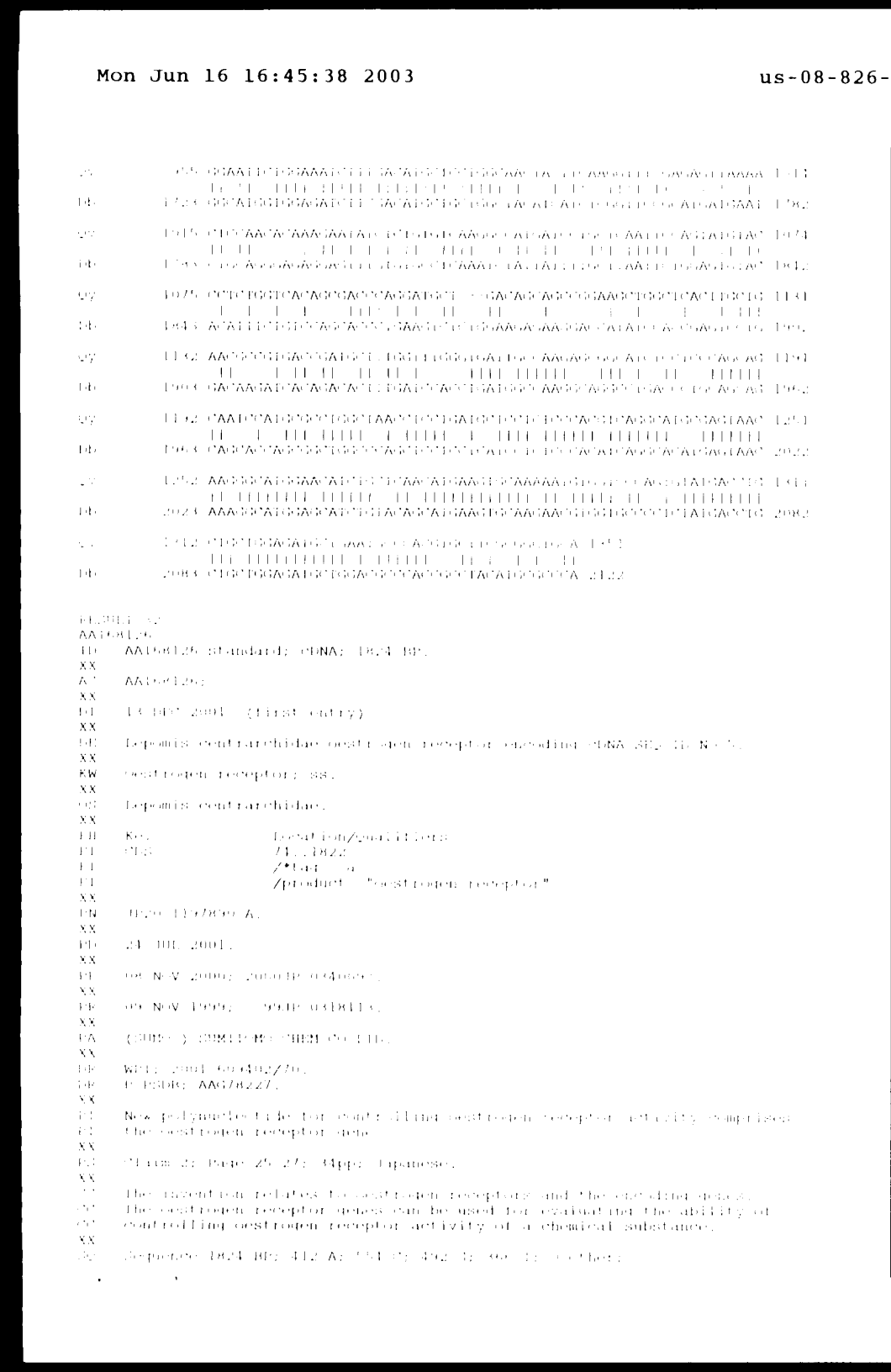
XX
AAX_[illegible]
XX
AC AAX_[illegible];
XX
DT [illegible]
XX
DE [illegible] progesterone receptor gene [illegible] MV* [illegible]
XX
KW tTA transactivator; tetracycline-controllable transactivator;
KW expression regulation; tet operator linked gene therapy;
KW [illegible] treatment; viral disease; TNF MV* [illegible];
KW [illegible] progesterone receptor; gs.
XX
OS [illegible] sp.
OC Eukaryota; Chordata; Vertebrata;
XX
RN [1] [illegible]-A.
XX
RA [illegible];
XX
RT [illegible];
XX
RL [illegible]
XX
RN [2] [illegible]
XX
RL [illegible]
XX
RN [3] [illegible]
XX
RL [illegible]
XX
FH Key Location/Qualifiers
FT source 1..[illegible]
FT /organism="[illegible]"
FT /db_xref="taxon:[illegible]"
FT gene [illegible]
FT /gene="MV*"
FT CDS [illegible]
FT /gene="MV*"
FT /product="progesterone receptor"
FT /protein_id="[illegible]"
FT /translation="[illegible]"
XX
SQ Sequence [illegible] AA; [illegible] CA; [illegible] G; [illegible] T; [illegible] other;

Query Match [illegible]; Score [illegible]; DB [illegible]; Length [illegible];
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

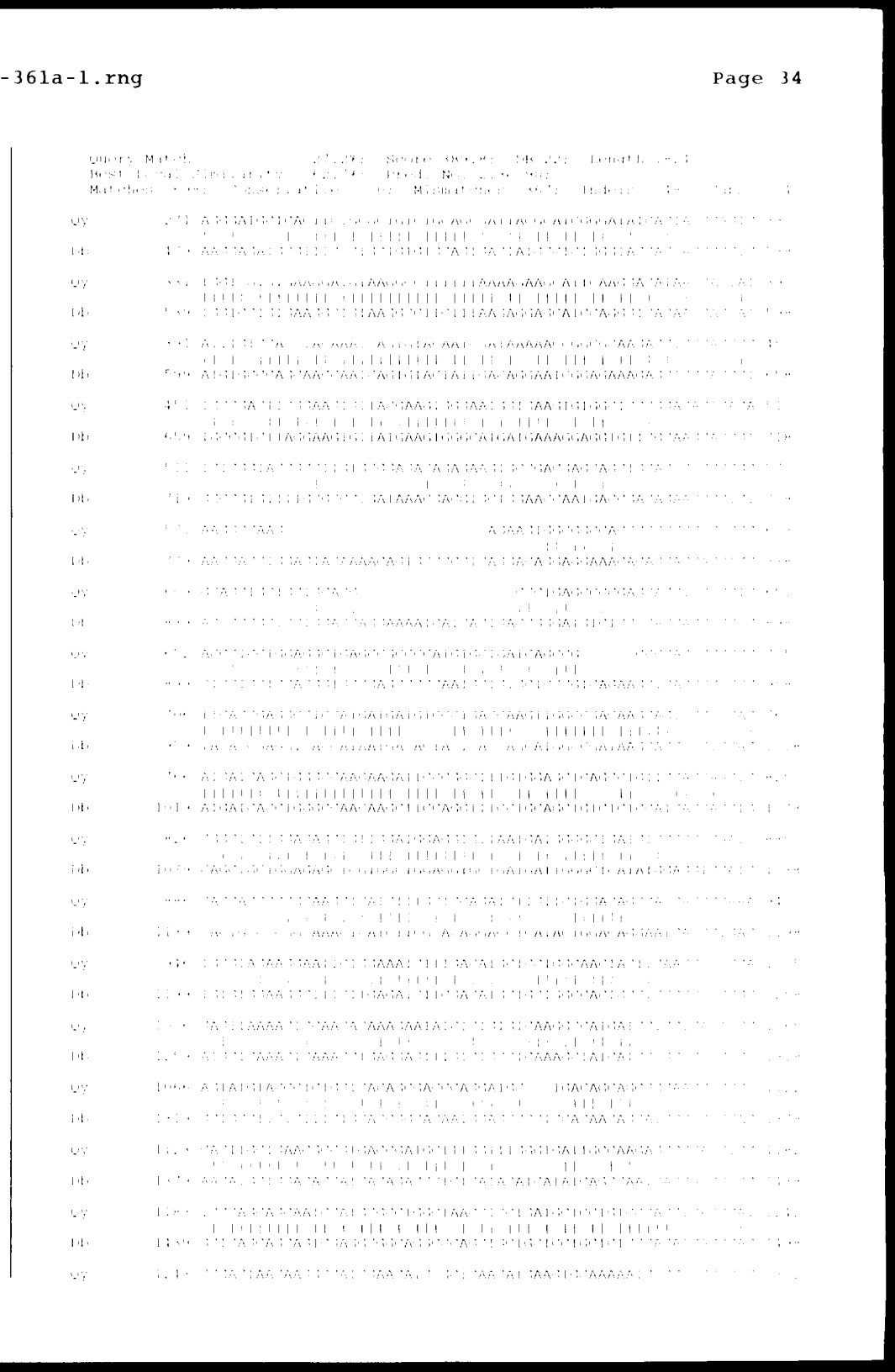


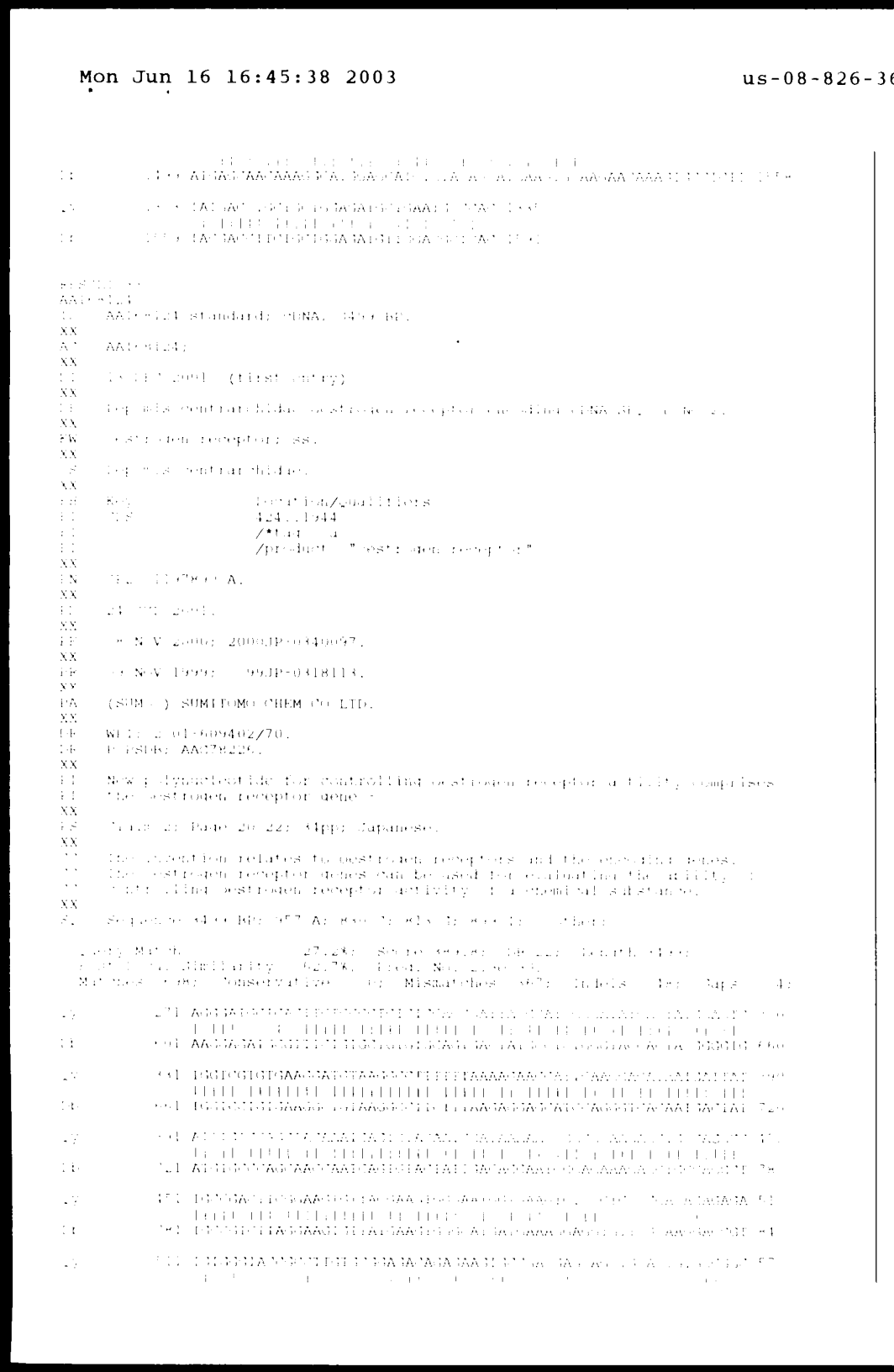
ID AAX61470 standard; DNA; 4665 BP.
XX
AC AAX61470;
XX
DT 16-APR-1999 (first entry)
XX
DE pTetMV4 rabbit progesterone fusion sequence.
XX
KW TTA transactivator; transgene; mouse; tet operator linked gene;
KW tetracycline controllable transactivator fusion protein;
KW transcriptional regulatory element; ds.
XX
OS Synthetic.
XX
RN [1]
RP 1-4665
RA
RD 12-JAN-1999.
XX
RF 07-JUN-1999; 99US-04481970.
XX
RF 07-JUN-1999; 99US-04481971.
RF 14-JUN-1999; 99US-00326327.
RF 14-JUN-1999; 99US-02604502.
XX
PA (PARE-) BASF AG.
XX
RT "Bujard H, Gossen M, Sillicl H, Vosa W;
XX
RL WPI; 1999-152108/13.
XX
PT Transgenic mouse - with tetracycline dependent gene suppression
XX
PS Disclosure; Fig 10; 79pp; English.
XX
CC This sequence represents a fusion gene between the the PhCMV* 1 DNA
CC sequence (containing the tet operator sequence and the hCMV minimal
CC promoter) and the rabbit progesterone receptor coding sequence. The
CC invention relates to a transgenic mouse having a transgene which
CC has the gene of the present sequence under the control of a tet
CC operator sequence, where the transgene comprises a
CC transcription activator element functional in cells of the
CC organism, linked to a polynucleotide sequence encoding a
CC tetracycline controllable transactivator fusion protein (TTA). The
CC protein comprises a tet repressor operably linked to a polypeptide which
CC directly or indirectly activates transcription of the tet operator linked
CC gene in eukaryotic cells. The tet operator linked gene encodes a
CC detectable and functional phenotype on the mouse when expression
CC of the gene. The transgene is expressed in cells of the mouse at a level
CC sufficient to produce amounts of the fusion protein that are sufficient
CC to activate transcription of the tet operator linked gene and in the
CC absence of tetracycline or a tetracycline analogue in the mouse the
CC fusion protein binds to the tet operator linked gene and activates
CC transcription of the tet operator linked gene such that the tet
CC operator linked gene is expressed at a level sufficient to produce the
CC detectable and functional phenotype on the mouse, where the level of
CC expression of tet operator linked gene can be down modulated by
CC administering tetracycline or a tetracycline analogue to the mouse. The
CC mice are useful as conditional knockout animals for studying human
CC diseases.
XX
Sq Sequence 4665 BP; 1274 A; 1078 C; 1178 G; 1135 T; 0 other;

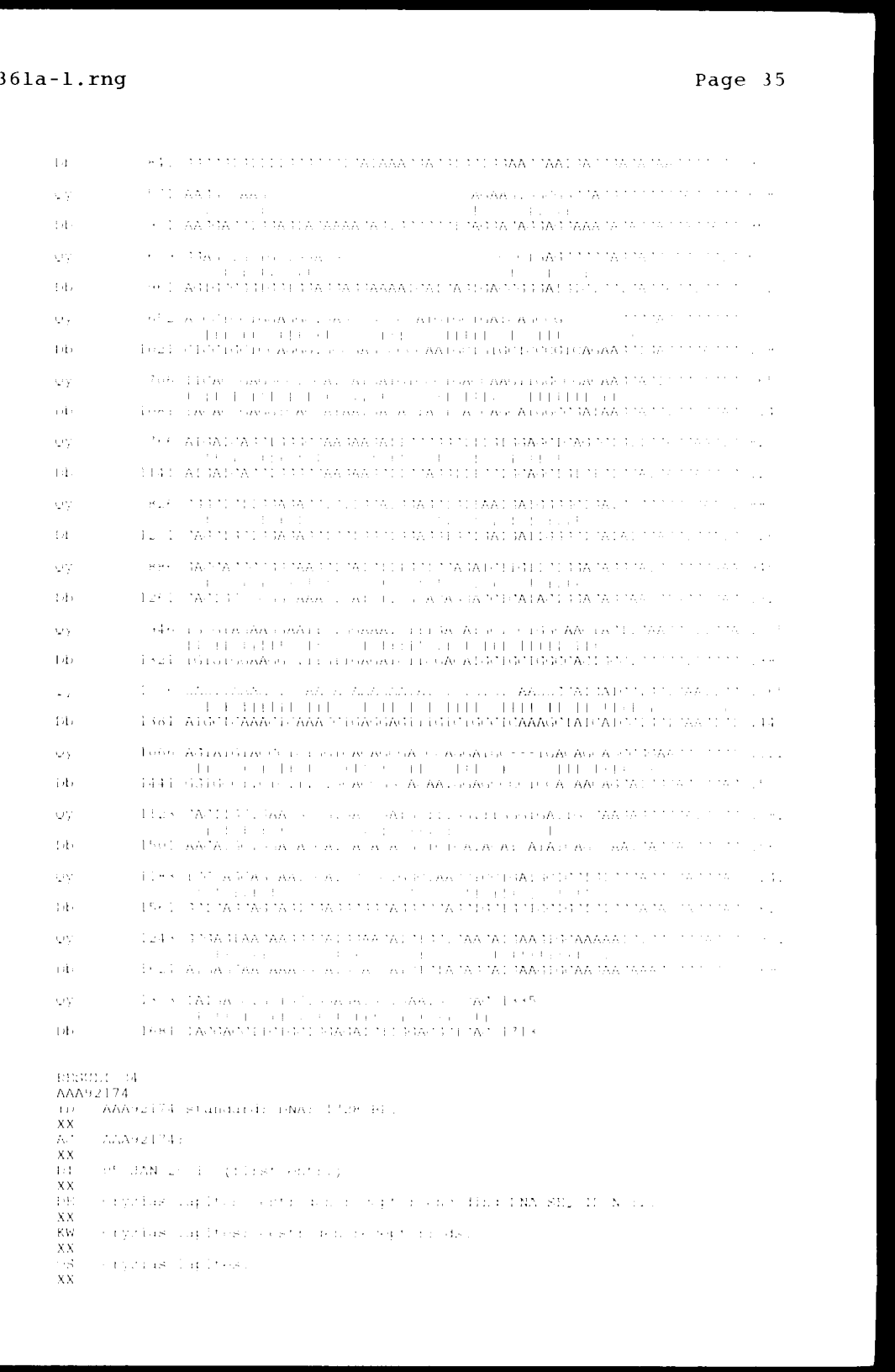




LOCUS
AA106126
DEFINITION standard; cDNA; 1824 BP.
XX
ACCESSION AA106126
XX
DT 19-DEC-2001 (first entry)
XX
DE Lepomis centrarchidae oestrogen receptor encoding cDNA SEQ ID NO 5.
XX
KW oestrogen receptor; ss.
XX
OS Lepomis centrarchidae.
XX
FH Key Location/qualifiers
FT misc_difference 71..1822
FT /label= a
FT /product= "oestrogen receptor"
XX
PN WO0177160-A.
XX
PD 21-JUN-2001.
XX
PF 08-NOV-2000; 2000WO-US030677.
XX
PR 09-NOV-1999; 99US-0438811.
XX
PA (SUBS-) SUBSTRATE GENETICS LTD.
XX
DR WPI; 2001-611822/70.
DR P-PSDB; AA078227.
XX
PT New polynucleotide for coding fish oestrogen receptor protein, useful for
PT the oestrogen receptor gene
XX
PS Claim 2; Page 25; 27pp; Japanese.
XX
CC The invention relates to oestrogen receptors and the encoding genes.
CC The oestrogen receptor genes can be used for evaluating the ability of
CC controlling oestrogen receptor activity of a chemical substance.
XX
SQ Sequence 1824 BP; 412 A; 551 C; 492 G; 369 T; 00 U; 0 Other;







```
PN   JP2000201680-A.
XX
PD   25-JUL-2000.
XX
PF   06-APR-1999;  99JP-0098[9].
XX
PR   16-NOV-1998;  98JP-0325467.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 2000-567760/51.
DR   P-PSDB; AAB20847.
XX
PT   An oestrogen receptor gene and its application
XX
PS   Claim 1; Page 11 14; 24pp; Japanese.
XX
CC   The present sequence encodes an oestrogen receptor derived from
CC   oryzias lapites. Also described are: (1) a vector comprising the
CC   oestrogen receptor gene; (2) a transformant prepared by introducing
CC   the oestrogen receptor gene or vector from (1) into a host cell;
CC   (3) a method for the preparation of an oestrogen receptor comprising
CC   culturing the transformant from (2) to produce the oestrogen receptor;
CC   and (4) a method for the evaluation of oestrogen receptor activating
CC   ability of a chemical substance in which the chemical substance is
CC   reacted with a transformant prepared by introducing a reporter gene
CC   connected downstream of a transcription controlling region containing
CC   an oestrogen response sequence and the above oestrogen receptor gene to
CC   an oestrogen nonendogenous host cell. The transformant can be used for
CC   the evaluation of oestrogen receptor activating ability of a chemical
CC   substance.
XX
SQ   Sequence 1728 BP; 478 A; 534 C; 497 G; xxxx T; 0 other;

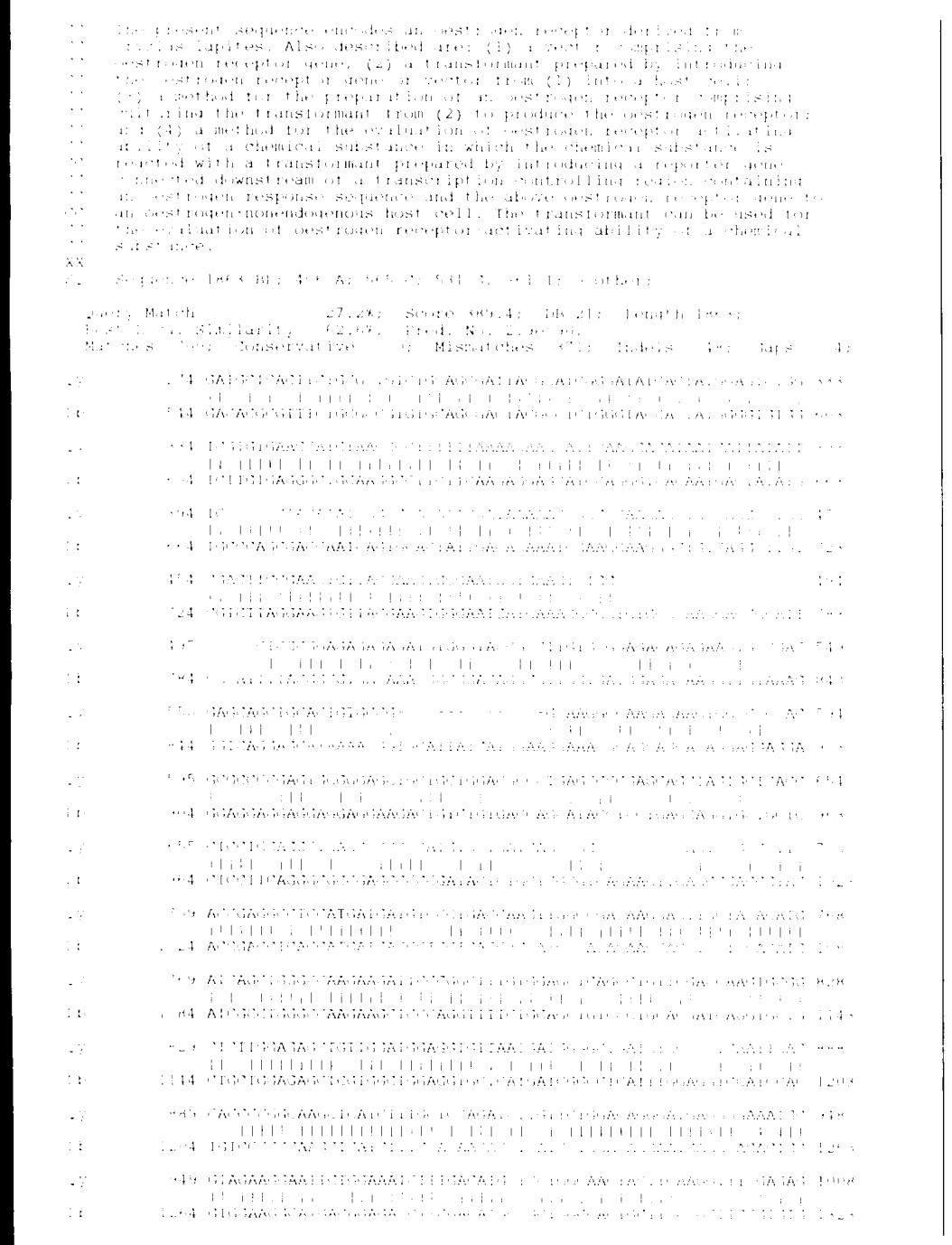
  Query Match             27.2%;  Score 995.4;  DB 21;  Length 1728;
  Best Local Similarity   67.6%;  Pred. No. 2.3e-66;
  Matches  700;  Conservative    0;  Mismatches  271;  Indels   48;  Gaps    4;

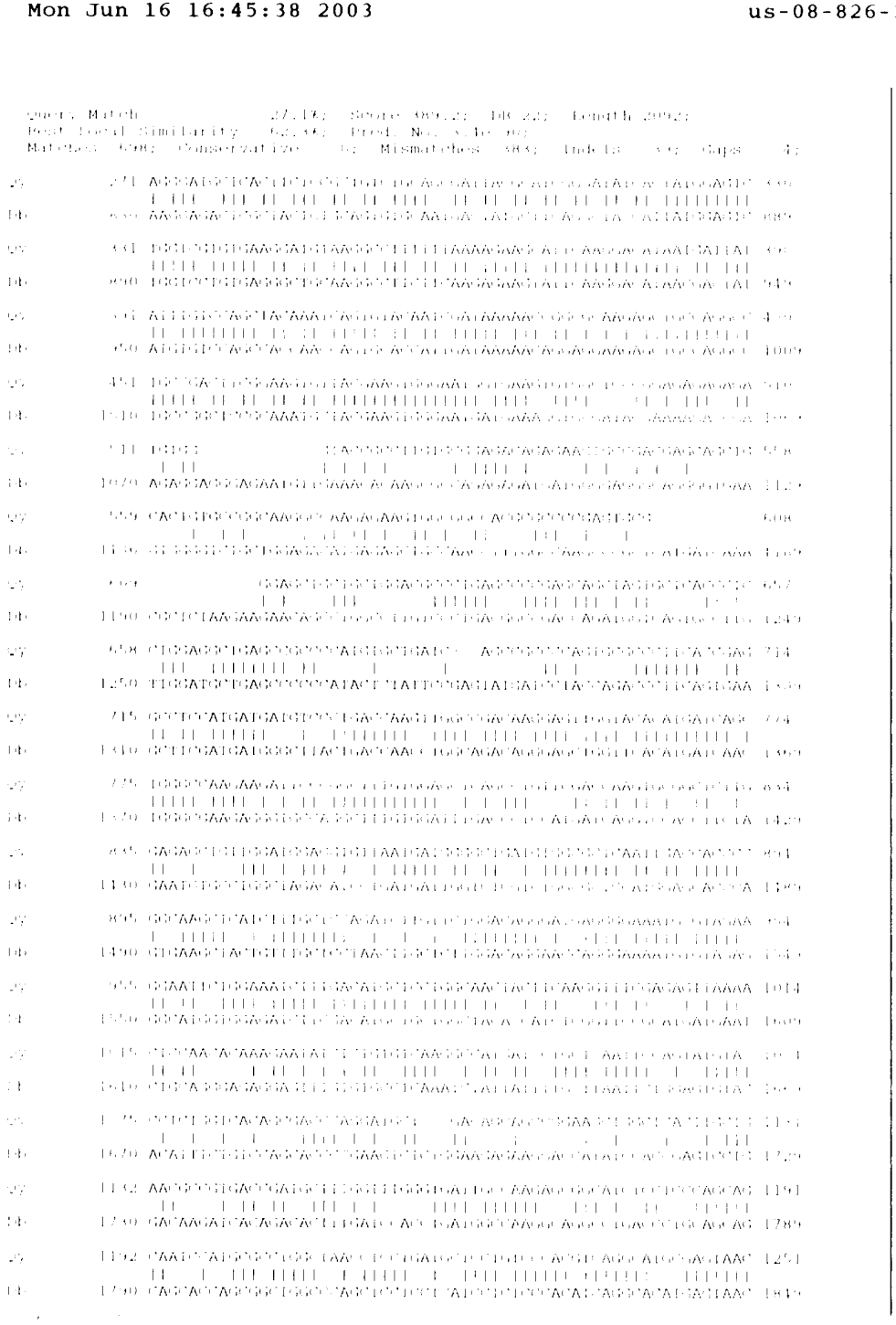
Qy         274 GAIGCTCACTTCTGCGAGGTCACAGCGATCTGCACGGATCATGCCGCCTGCG 333
               || | | ||||||| | |||||||| |||| || || || |||| || || |||| ||||
Db         100 GACACGCGTTTCTGCGAGGTCGCGGCGATCTGCGCGGTCATGCCGCCTGCG 153

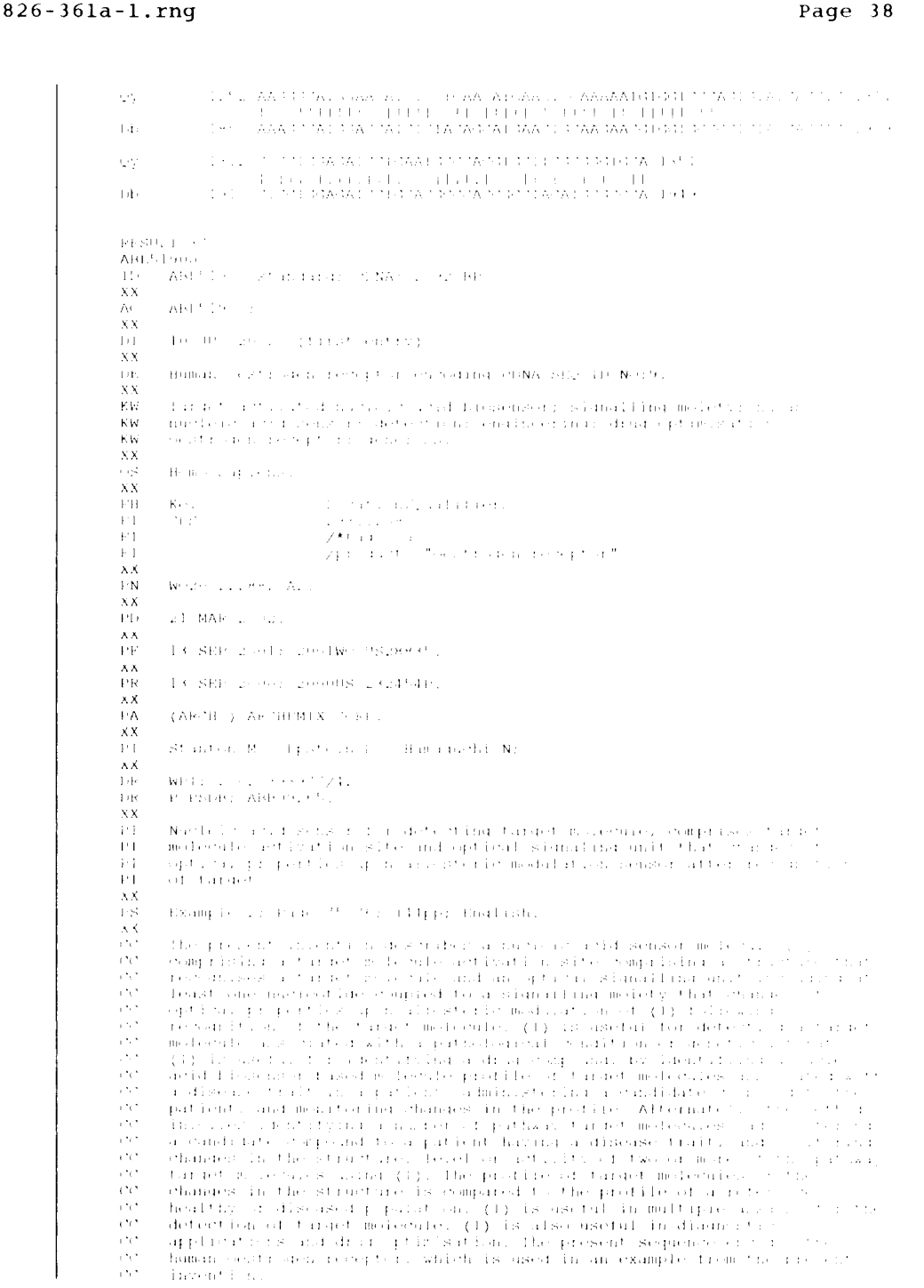
Qy         334 TCTGCGAAGGATGTAAGGTCTTTTTAAAAAAGTCATCAAGTATATATTT 383
               || || || || || || ||| |||| || || || |||| || || |||| ||||
Db         154 TCTGCGAGGGCTGCAAGGTCTTCTACAAGAGGACCATCCAGGGCATATTT 203

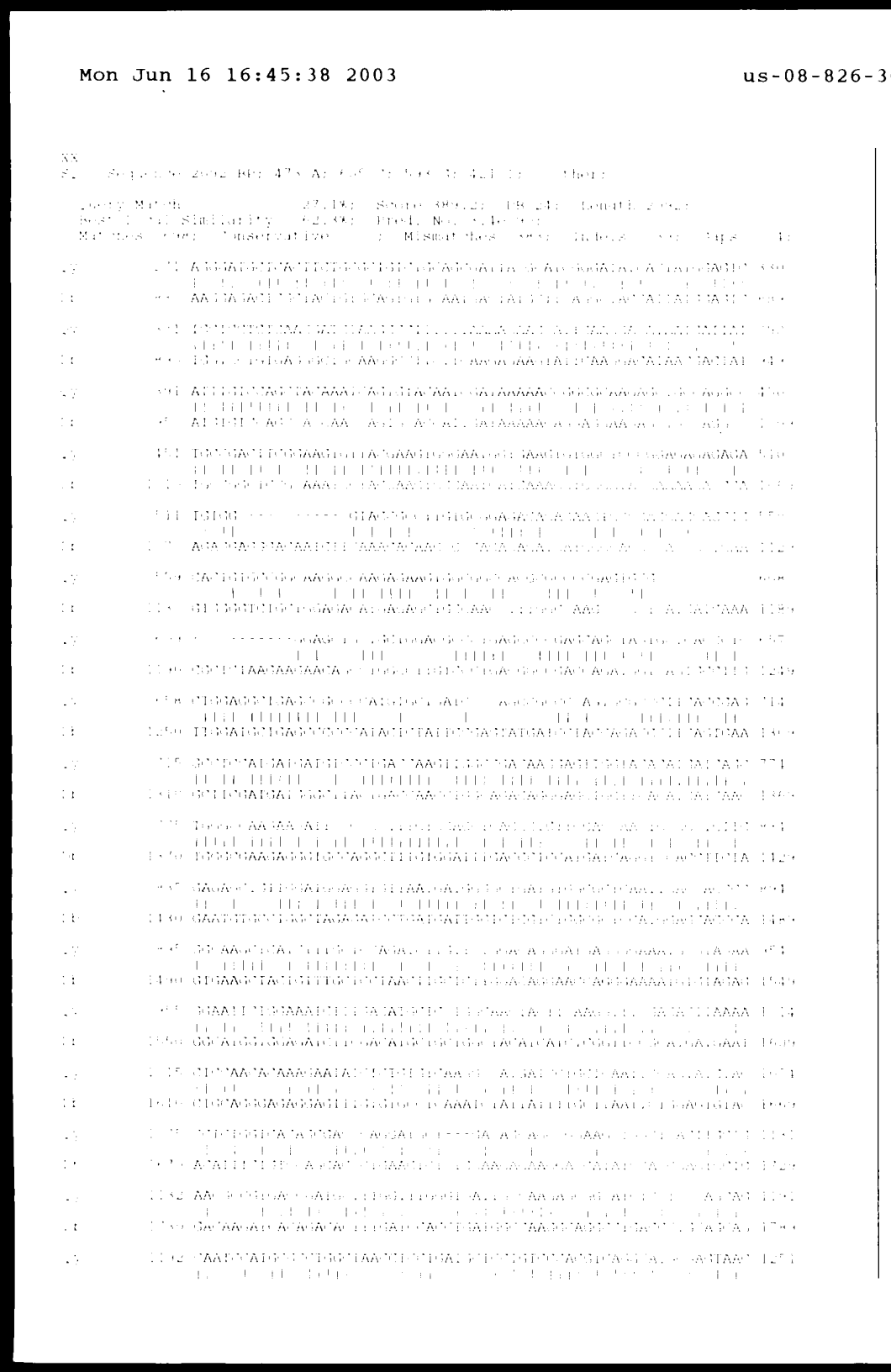
Qy         384 TGTCCAGCTACAAATGCATCTTACAATCGATAAAAA-CCCCCAAGACCTCCAGTCTC 442
               || || || || |||||| | || || || | | || | | || |||| ||
Db         204 TGCCAGCTACAAATGCACCTACGACTGCGTCATCGACAAGCACCAGCGCAAGAC 259

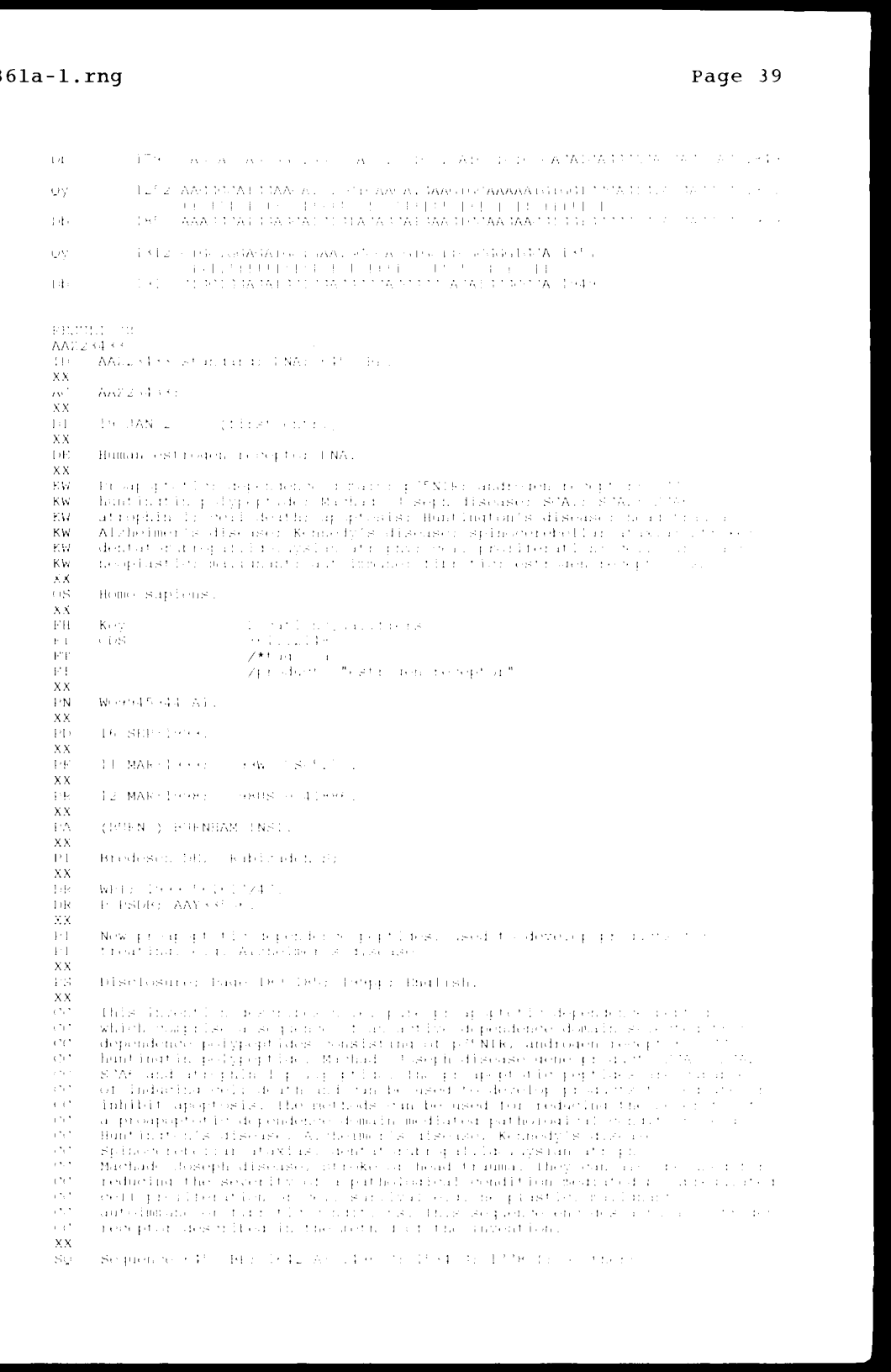
Qy         443 CGACTCCGGAAGGTGCGAGGCGTGCCGCCTCCAAGAAGTGCTTCGTCGTCGGC 492
               || || | | |
```

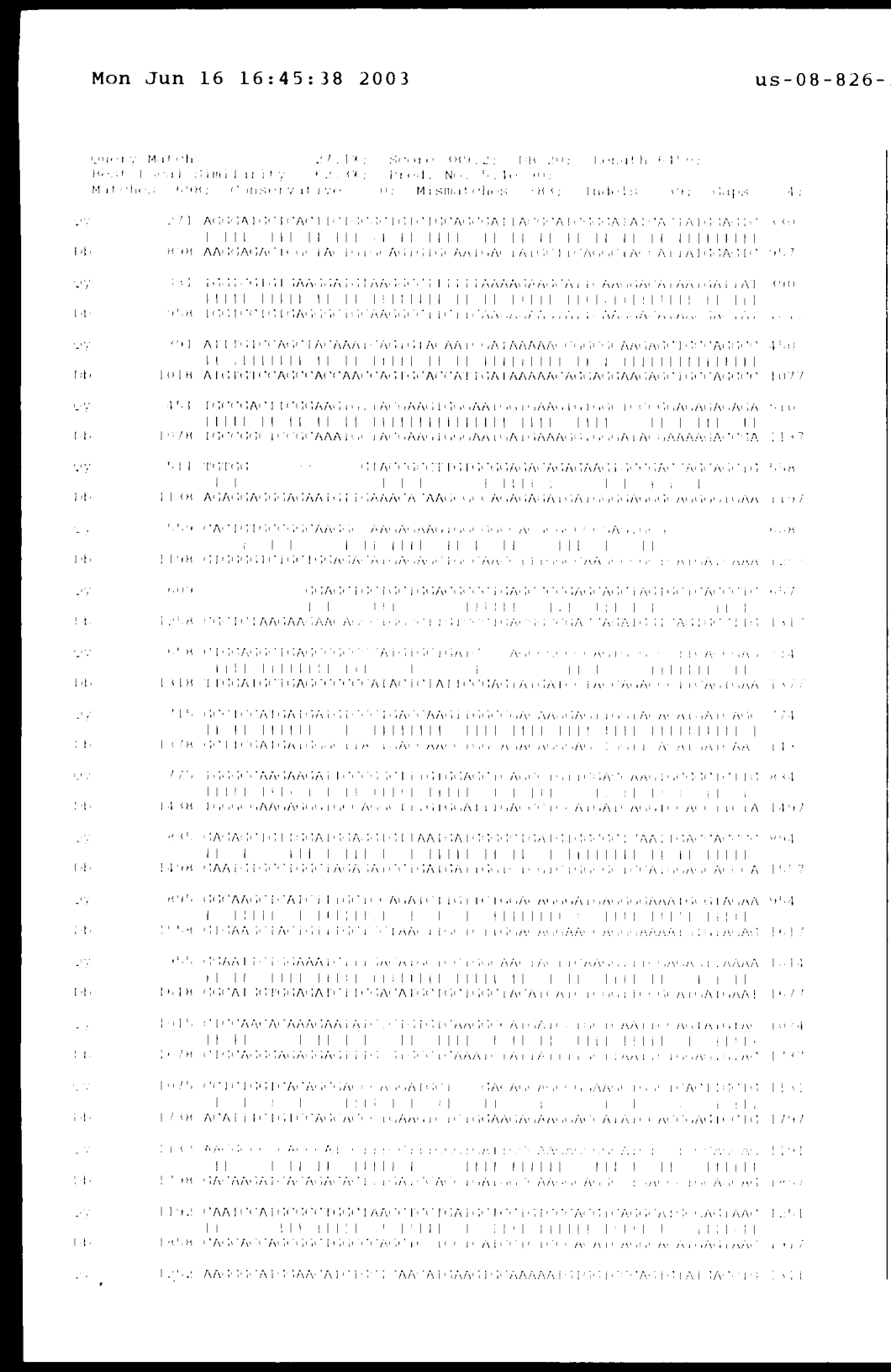


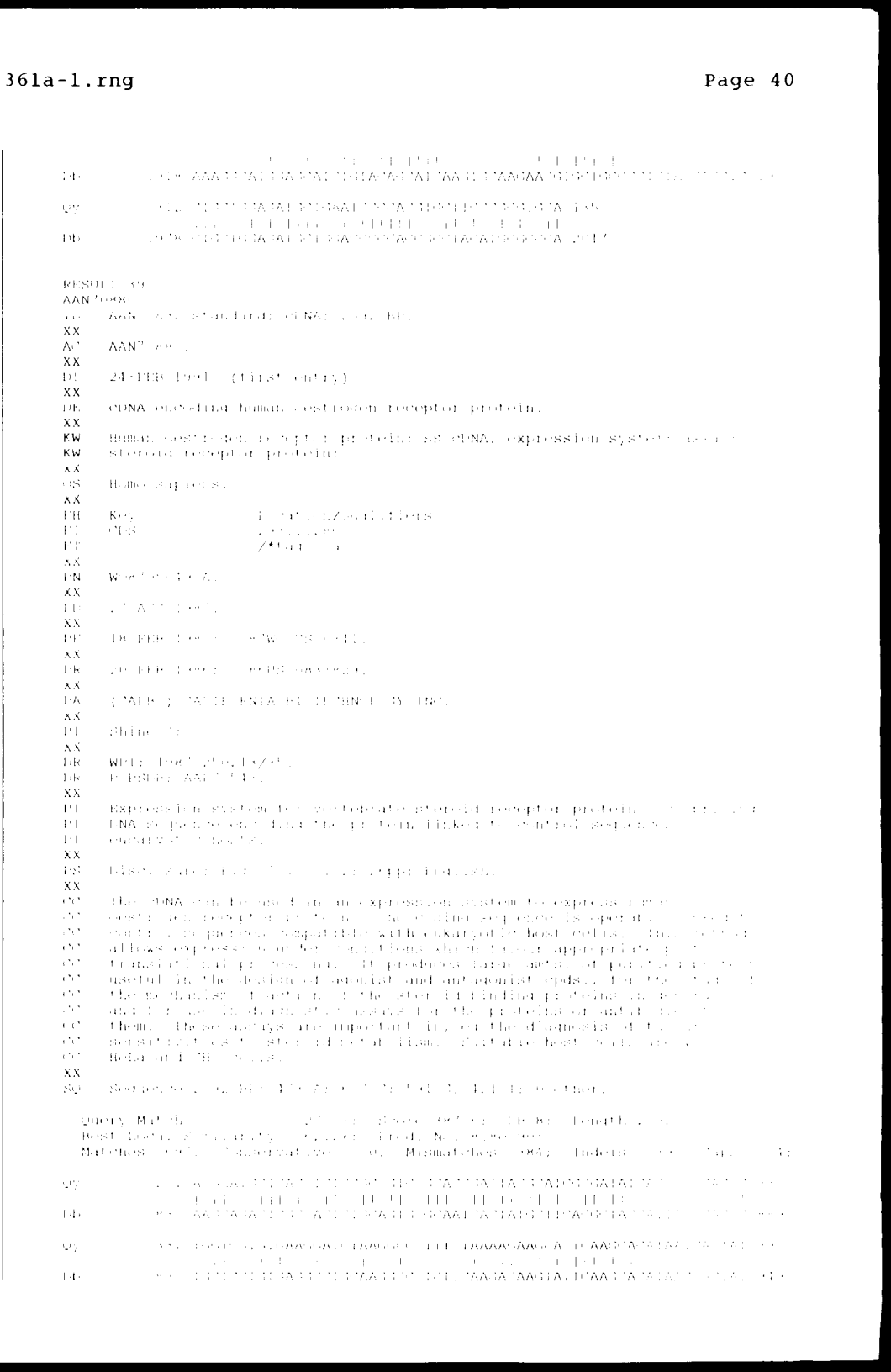


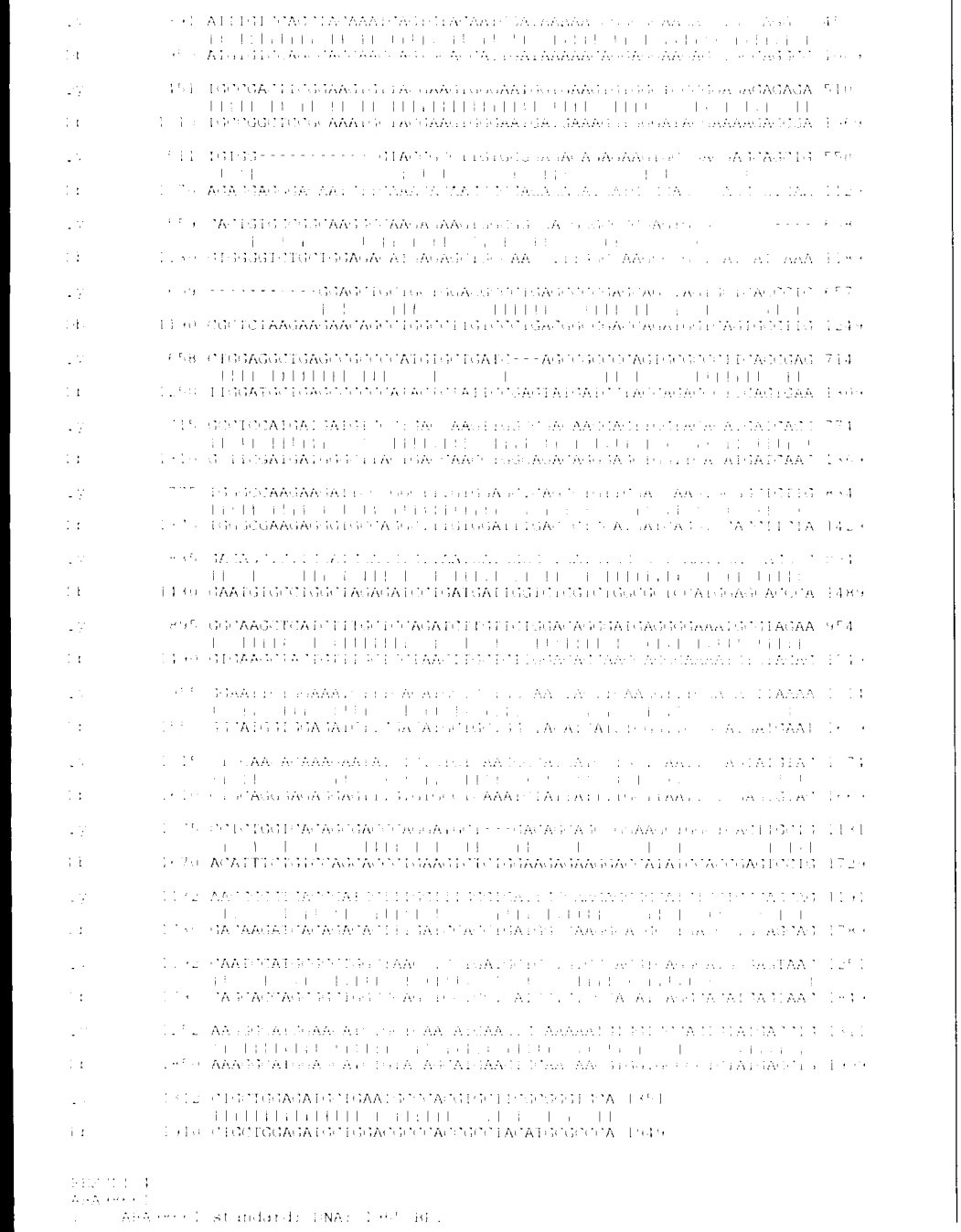


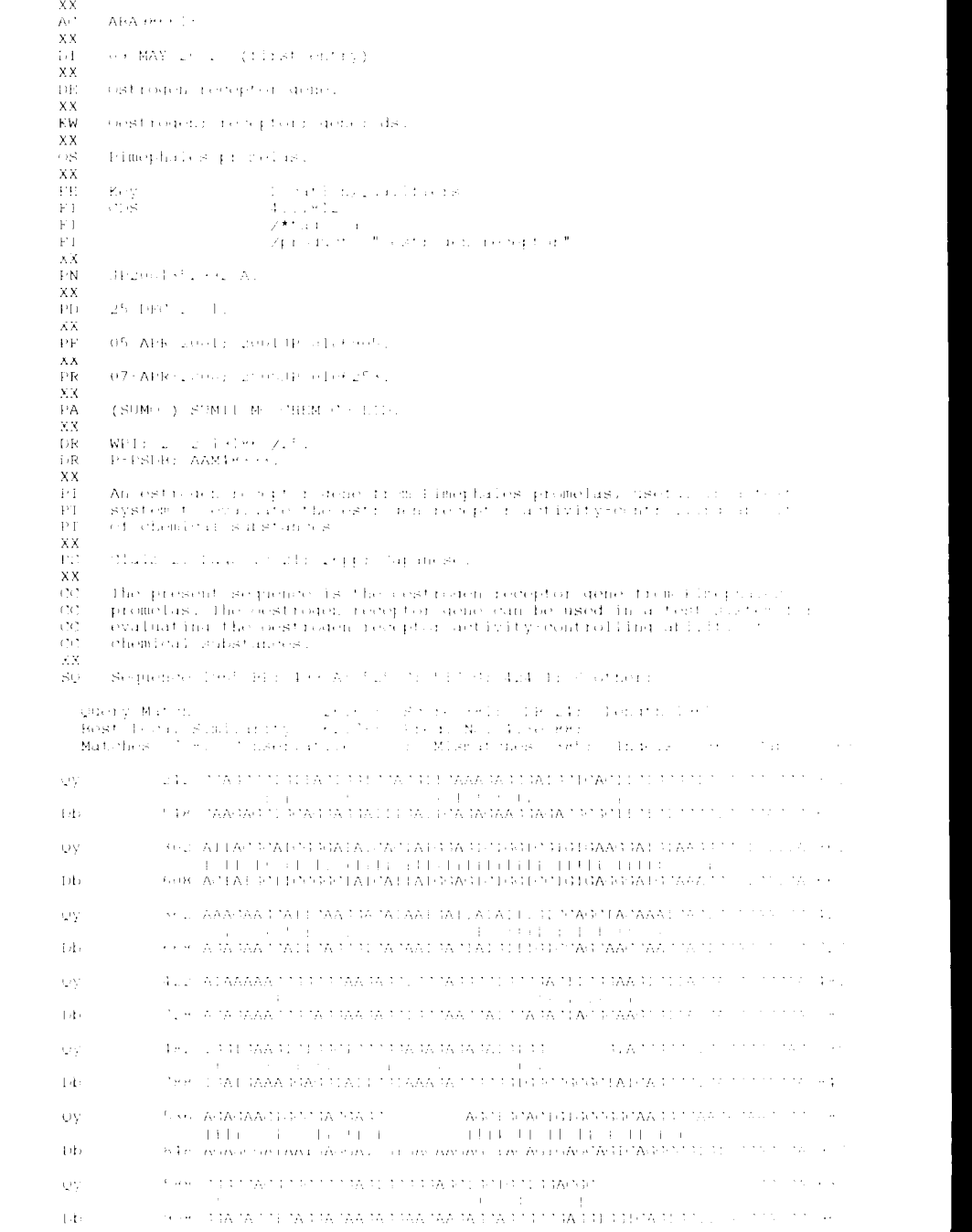


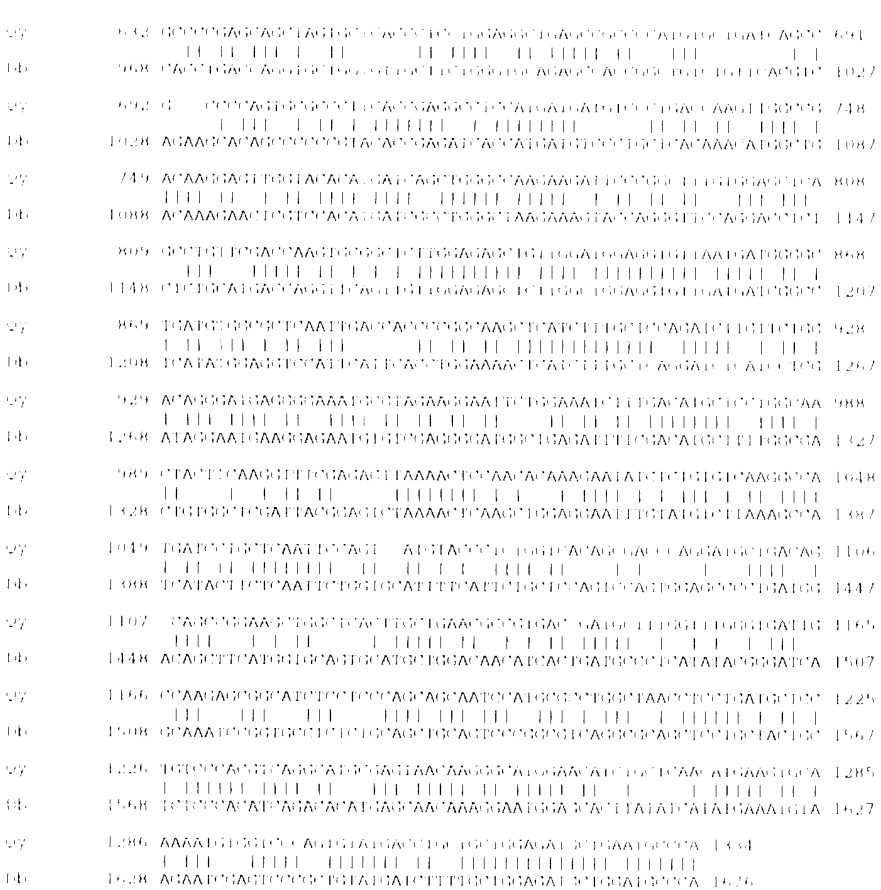












Search Completed: June 16, 2003, 00:47:43
Job Time : 339.789 Secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

DM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 09:20:22 ; Search time 89.375 Seconds
(without alignments)
6512.699 Million cell updates/sec

Title: US-08-826-361A-24
Perfect score: 1898
Sequence: 1 CACGAATCTTTGAGAACATT.....TGAATCGAGCGTTTCCACAC 1898

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1635.6	86.2	1647	4	US-09-139-617-2	Sequence 2, Appli
2	1635.6	86.2	1647	4	US-09-561-741A-2	Sequence 2, Appli
3	1454.6	76.6	1460	2	US-08-836-620A-4	Sequence 4, Appli
4	1239.6	65.3	2568	2	US-08-836-620A-1	Sequence 1, Appli
5	1099.6	57.9	1458	2	US-08-836-620A-6	Sequence 6, Appli
6	390.8	20.6	4963	1	US-08-076-726-16	Sequence 16, Appl
7	390.8	20.6	4963	1	US-08-260-452-9	Sequence 9, Appli
8	390.8	20.6	4963	2	US-08-481-970-9	Sequence 9, Appli
9	390.8	20.6	4963	2	US-08-897-719-9	Sequence 9, Appli
10	390.8	20.6	4963	4	US-09-163-269-9	Sequence 9, Appli
11	389.2	20.5	6450	4	US-09-041-886-34	Sequence 34, Appl
12	389.2	20.5	6450	4	US-08-453-908-1	Sequence 1, Appli
13	285.6	15.0	1956	3	US-08-693-940-2	Sequence 2, Appli
14	285.6	15.0	1956	4	US-09-566-669-2	Sequence 2, Appli
15	285.6	15.0	2322	3	US-08-564-264-2	Sequence 2, Appli
16	114.8	6.0	1615	3	US-09-040-508-1	Sequence 1, Appli
17	114.8	6.0	1615	4	US-09-500-654-1	Sequence 1, Appli
18	95.2	5.0	3715	4	US-09-041-886-10	Sequence 10, Appl
19	89	4.7	6244	1	US-08-076-726-15	Sequence 15, Appl
20	89	4.7	6244	1	US-08-260-452-8	Sequence 8, Appli
21	89	4.7	6244	2	US-08-481-970-8	Sequence 8, Appli
22	89	4.7	6244	2	US-08-897-719-8	Sequence 8, Appli
23	89	4.7	6244	4	US-09-163-269-8	Sequence 8, Appli
24	84.8	4.5	6177	4	US-08-479-918E-1	Sequence 1, Appli
25	84.8	4.5	7257	4	US-09-091-042A-1	Sequence 1, Appli
26	81	4.3	3014	1	US-08-629-939-1	Sequence 1, Appli
27	81	4.3	3014	1	US-08-759-873-1	Sequence 1, Appli
28	80.8	4.3	1576	6	5260432-1	Patent No. 5260432
29	78	4.1	2489	6	5223606-1	Patent No. 5223606
30	74.4	3.9	1797	4	US-09-277-078-1	Sequence 1, Appli
31	74.4	3.9	2658	2	US-08-592-383-3	Sequence 3, Appli
32	74.4	3.9	2928	2	US-08-095-728B-3	Sequence 3, Appli
33	74.4	3.9	2928	5	PCT-US92-02320A-3	Sequence 3, Appli
34	74.4	3.9	2940	2	US-08-592-383-1	Sequence 1, Appli
35	74.4	3.9	2940	6	5171671-1	Patent No. 5171671
36	74.4	3.9	3036	1	US-08-306-691B-52	Sequence 52, Appl
37	74.4	3.9	3036	2	US-08-095-728B-1	Sequence 1, Appli
38	74.4	3.9	3036	5	PCT-US92-02320A-1	Sequence 1, Appli
39	74.4	3.9	3460	1	US-08-312-312A-1	Sequence 1, Appli
40	74.4	3.9	3511	3	US-08-892-747-13	Sequence 13, Appl
41	73.2	3.9	201	2	US-08-592-383-6	Sequence 6, Appli
42	72.8	3.8	2067	4	US-09-106-194-11	Sequence 11, Appl
43	71.2	3.8	387	1	US-08-087-151-1	Sequence 1, Appli
44	71.2	3.8	387	1	US-08-561-072-1	Sequence 1, Appli
45	71.2	3.8	1662	1	US-08-336-408B-5	Sequence 5, Appli
46	71.2	3.8	1662	5	PCT-US91-00399-5	Sequence 5, Appli
47	66.4	3.5	1641	3	US-08-705-771-9	Sequence 9, Appli
48	65.4	3.4	2233	1	US-08-496-631-1	Sequence 1, Appli
49	65.2	3.4	1952	1	US-08-333-358-1	Sequence 1, Appli
50	65.2	3.4	1952	1	US-08-463-694-1	Sequence 1, Appli
51	65.2	3.4	1952	1	US-08-694-501-1	Sequence 1, Appli
52	64.8	3.4	2095	1	US-08-336-408B-3	Sequence 3, Appli
53	64.8	3.4	2095	5	PCT-US91-00399-3	Sequence 3, Appli
54	64.8	3.4	2711	4	US-08-891-298-2	Sequence 2, Appli
55	64	3.4	1245	2	US-09-142-619-3	Sequence 3, Appli
56	64	3.4	1245	3	US-09-282-803B-3	Sequence 3, Appli
57	64	3.4	1245	4	US-09-510-654-3	Sequence 3, Appli
58	64	3.4	2338	2	US-09-142-619-7	Sequence 7, Appli
59	64	3.4	2338	3	US-09-282-803B-7	Sequence 7, Appli
60	64	3.4	2338	4	US-09-510-654-7	Sequence 7, Appli
61	64	3.4	3027	2	US-09-142-619-9	Sequence 9, Appli
62	64	3.4	3027	3	US-09-282-803B-9	Sequence 9, Appli
63	64	3.4	3027	4	US-09-510-654-9	Sequence 9, Appli
64	64	3.4	3115	2	US-09-142-619-1	Sequence 1, Appli
65	64	3.4	3115	3	US-09-282-803B-1	Sequence 1, Appli
66	64	3.4	3115	4	US-09-510-654-1	Sequence 1, Appli
67	64	3.4	3251	2	US-09-142-619-5	Sequence 5, Appli
68	64	3.4	3251	3	US-09-282-803B-5	Sequence 5, Appli
69	64	3.4	3251	4	US-09-510-654-5	Sequence 5, Appli
70	63.6	3.4	2094	4	US-09-106-194-1	Sequence 1, Appli
71	62	3.3	2130	1	US-07-952-800-1	Sequence 1, Appli
72	61.2	3.2	2468	1	US-08-333-358-11	Sequence 11, Appl
73	61.2	3.2	2468	1	US-08-463-694-11	Sequence 11, Appl
74	61.2	3.2	2468	1	US-08-694-501-11	Sequence 11, Appl
75	60.8	3.2	1796	4	US-09-255-392-1	Sequence 1, Appli
76	60.8	3.2	2005	2	US-08-484-200-1	Sequence 1, Appli
77	60.8	3.2	2005	2	US-08-477-494-1	Sequence 1, Appli
78	60.8	3.2	2005	3	US-08-465-375-1	Sequence 1, Appli
79	60.8	3.2	2005	4	US-09-788-070-1	Sequence 1, Appli
80	60.4	3.2	1866	1	US-08-336-408B-1	Sequence 1, Appli
81	60.4	3.2	1866	5	PCT-US91-00399-1	Sequence 1, Appli
82	59.2	3.1	2204	1	US-07-952-800-3	Sequence 3, Appli
83	59.2	3.1	2219	1	US-08-336-408B-7	Sequence 7, Appli
84	59.2	3.1	2219	5	PCT-US91-00399-7	Sequence 7, Appli
85	59	3.1	1677	2	US-08-372-652-14	Sequence 14, Appl
86	59	3.1	1677	5	PCT-US95-16311-14	Sequence 14, Appl
87	59	3.1	1787	2	US-08-372-652-6	Sequence 6, Appli
88	59	3.1	1787	5	PCT-US95-16311-6	Sequence 6, Appli
89	58.8	3.1	1748	1	US-08-255-471-8	Sequence 8, Appli
90	57	3.0	1441	4	US-08-927-219-78	Sequence 78, Appl
91	56.8	3.0	1934	4	US-08-653-648A-2	Sequence 2, Appli
92	56.8	3.0	1934	4	US-08-653-648A-9	Sequence 9, Appli
93	56.8	3.0	2463	4	US-08-653-648A-10	Sequence 10, Appl
94	56.8	3.0	2464	4	US-08-653-648A-3	Sequence 3, Appli
95	56.8	3.0	2745	4	US-08-653-648A-4	Sequence 4, Appli
96	54.4	2.9	1434	4	US-09-128-142-1	Sequence 1, Appli
97	54.4	2.9	1519	4	US-09-128-142-2	Sequence 2, Appli
98	54.4	2.9	1608	3	US-09-484-345-3	Sequence 3, Appli
99	54.4	2.9	1679	4	US-09-514-247A-5	Sequence 5, Appli
100	54.4	2.9	1844	4	US-08-134-557D-1	Sequence 1, Appli

101	54.6	2.8	2304	1	US-08-464-266-1	Sequence 1, Appli
102	54.6	2.8	2304	1	US-08-464-272-1	Sequence 1, Appli
103	54.6	2.8	2304	4	US-08-464-514-1	Sequence 1, Appli
104	54.6	2.8	2304	4	US-08-486-403-1	Sequence 1, Appli
105	52.2	2.8	1758	1	US-08-078-222B-1	Sequence 1, Appli
106	52.2	2.8	1758	2	US-08-661-330A-1	Sequence 1, Appli
107	52.2	2.8	1758	3	US-09-048-217A-1	Sequence 1, Appli
108	50	2.6	1399	1	US-07-747-746B-5	Sequence 5, Appli
109	50	2.6	2043	1	US-07-747-746B-6	Sequence 6, Appli
110	47.6	2.5	2315	1	US-08-333-358-13	Sequence 13, Appl
111	47.6	2.5	2315	1	US-08-463-694-13	Sequence 13, Appl
112	47.6	2.5	2315	1	US-08-694-501-13	Sequence 13, Appl
113	47.2	2.5	1450	1	US-08-459-489-1	Sequence 1, Appli
114	47.2	2.5	1450	1	US-08-458-686-1	Sequence 1, Appli
115	47.2	2.5	1450	1	US-07-843-350C-1	Sequence 1, Appli
116	47.2	2.5	1450	5	PCT-US93-01559-1	Sequence 1, Appli
117	46.6	2.5	1649	2	US-08-466-120-1	Sequence 1, Appli
118	46.6	2.5	1649	5	PCT-US94-07266-1	Sequence 1, Appli
119	46.6	2.5	1659	1	US-08-333-358-7	Sequence 7, Appli
120	46.6	2.5	1659	1	US-08-463-694-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-09-139-617-2
; Sequence 2, Application US/09139617
; Patent No. 6222015
; GENERAL INFORMATION:
;  APPLICANT: WILKINSON, HILARY
;  TITLE OF INVENTION: ESTROGEN RECEPTOR
;  FILE REFERENCE: 20047Y
;  CURRENT APPLICATION NUMBER: US/09/139,617
;  CURRENT FILING DATE: 1998-08-25
;  EARLIER APPLICATION NUMBER: 60/058,271
;  EARLIER FILING DATE: 1997-09-08
;  EARLIER APPLICATION NUMBER: 60/060,520
;  EARLIER FILING DATE: 1997-09-30
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1647
;   TYPE: DNA
;   ORGANISM: HUMAN
US-09-139-617-2

  Query Match             86.2%;  Score 1635.6;  DB 4;  Length 1647;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1638;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAA 87
               | ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           6 CTTTGTAGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAA 65

Qy          88 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 147
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          66 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 125

Qy         148 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 207
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         126 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 185

Qy         208 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 267
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         186 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 245

Qy         268 CTTGGAAGGTGGGCCTCGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGG 327
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         246 CTTGGAAGGTGGGCCTCGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGG 305

Qy         328 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 387
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         306 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 365

Qy         388 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 447
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         366 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 425

Qy         448 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 507
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         426 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 485

Qy         508 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 567
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         486 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 545

Qy         568 GTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATAT 627
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         546 GTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATAT 605

Qy         628 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCCAAGAGCTGCCAGGCCTG 687
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         606 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCCAAGAGCTGCCAGGCCTG 665

Qy         688 CCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATG 747
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         666 CCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATG 725

Qy         748 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 807
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         726 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 785

Qy         808 GGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAG 867
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         786 GGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAG 845

Qy         868 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 927
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         846 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 905

Qy         928 CCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 987
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         906 CCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 965

Qy         988 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1047
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         966 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1025

Qy        1048 GTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1107
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1026 GTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1085

Qy        1108 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1167
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1086 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1145

Qy        1168 GGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1227
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1146 GGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1205

Qy        1228 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGAT 1287
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1206 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGAT 1265

Qy        1288 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1347
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1266 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1325

Qy        1348 GAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1407
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1326 GAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1385

Qy        1408 CGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1467
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1386 CGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1445
```

```
Qy       1468 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1505

Qy       1528 GGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACCGTCTTCGCGGGGTGCAA 1587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 GGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACCGTCTTCGCGGGGTGCAA 1565

Qy       1588 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGGACAGTAAAAGCAAACAGGGTGT 1647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGGACAGTAAAAGCAAAGAGGGTGT 1625

Qy       1648 CCAGAACCCACAGTCTCAGTGA 1669
             ||||||||||||||||||||||
Db       1626 CCAGAACCCACAGTCTCAGTGA 1647


RESULT 2
US-09-561-741A-2
; Sequence 2, Application US/09561741A
; Patent No. 6458551
; GENERAL INFORMATION:
;  APPLICANT: WILKINSON, HILARY
;  TITLE OF INVENTION: ESTROGEN RECEPTOR
;  FILE REFERENCE: 20047Y
;  CURRENT APPLICATION NUMBER: US/09/561,741A
;  CURRENT FILING DATE:  2000-04-26
;  PRIOR APPLICATION NUMBER: 09/139,617
;  PRIOR FILING DATE: 1998-08-25
;  PRIOR APPLICATION NUMBER: 60/058,271
;  PRIOR FILING DATE: 1997-09-08
;  PRIOR APPLICATION NUMBER: 60/060,520
;  PRIOR FILING DATE: 1997-09-30
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1647
;   TYPE: DNA
;   ORGANISM: HUMAN
US-09-561-741A-2

  Query Match             86.2%;  Score 1635.6;  DB 4;  Length 1647;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1638;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         28 CCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAA 87
              | ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 CTTTGTAGCCTCTTCTTGCAAGGTGTTTTCTCAGCTGTTATCTCAAGACATGGATATAAA 65

Qy         88 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 AAACTCACCATCTAGCCTTAATTCTCCTTCCTCCTACAACTGCAGTCAATCCATCTTACC 125

Qy        148 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 CCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAGCCACCATGAATATCC 185

Qy        208 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAA 245

Qy        268 CTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGG 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 CTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGG 305

Qy        328 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAA 365

Qy        388 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACT 425

Qy        448 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAG 485

Qy        508 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 GGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTG 545

Qy        568 GTCGTGTGAAGGATGTAAGGCCTTTTTTAAAACAAGCATTCAAGGACATAATGATTATAT 627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 GTCGTGTGAAGGATGTAAGGCCTTTTTTAAAACAAGCATTCAAGGACATAATGATTATAT 605

Qy        628 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTG 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 TTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTG 665

Qy        688 CCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATG 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 CCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATG 725

Qy        748 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 TGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAA 785

Qy        808 GGCCAAGAGAAGTGGCGGCCAGCGCCCCGAGTGCGGGAGCTGCTGTGGACGCCCTGAG 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 GGCCAAGAGAAGTGGCGGCCAGCGCCCCGAGTGCGGGAGCTGCTGTGGACGCCCTGAG 845

Qy        868 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 CCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCG 905

Qy        928 CCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 CCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAA 965

Qy        988 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 GGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCT 1025

Qy       1048 GTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 GTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGAT 1085

Qy       1108 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 GTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAG 1145

Qy       1168 GGATCAGGGGAAATGGGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 GGATGAGGGGAAATGGGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTAC 1205

Qy       1228 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAACAATATCTCTGTGTCAAGGCCATGAT 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 TTCAAGGTTTCGAGAGTTAAAACTCCAACACAAACAATATCTCTGTGTCAAGGCCATGAT 1265

Qy       1288 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 CCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCG 1325

Qy       1348 GAAGCTGGCTCACTTGCTGAACGCCCTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 GAAGCTGGCTCACTTGCTGAACGCCCTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAG 1385

Qy       1408 CGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 CGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCA 1445

Qy       1468 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 CGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGT 1505
```

```
Qy       1528 GGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAA 1587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 GGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAA 1565

Qy       1588 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTC 1647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 GTCCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTC 1625

Qy       1648 CCAGAACCCACAGTCTCAGTGA 1669
             |||||||||||||||||||||||
Db       1626 CCAGAACCCACAGTCTCAGTGA 1647


RESULT 3
US-08-826-620A-4
; Sequence 4, Application US/08826620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/826,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1460 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-826-620A-4

  Query Match             76.6%;  Score 1453.6;  DB 2;  Length 1460;
  Best Local Similarity   99.7%;  Prod. No. 0;
  Matches 1456;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        210 CCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 269
              | ||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTATGACATTCTACAGTCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACT 60

Qy        270 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGC 120

Qy        330 ACCTTTCTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTTCTCTCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGA 180

Qy        390 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGA 240

Qy        450 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGG 300

Qy        510 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGT 360

Qy        570 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTT 420

Qy        630 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCC 480

Qy        690 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGGAGAGAGATGTG 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGGAGAGAGATGTG 540

Qy        750 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGG 600

Qy        810 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCC 660

Qy        870 CCGACAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGACAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCC 720

Qy        930 CCACTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCACTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGG 780

Qy        990 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGT 840

Qy       1050 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 1109
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGT 900

Qy       1110 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 1169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGG 960

Qy       1170 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTT 1229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTT 1020

Qy       1230 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCC 1080

Qy       1290 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGA 1140

Qy       1350 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCG 1200

Qy       1410 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACG 1260

Qy       1470 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGG 1320

Qy       1530 TGCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1589
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGT 1380

Qy       1590 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1649
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTCCATCACGGGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCC 1440

Qy       1650 AGAACCCACAGTCTCAGTGA 1669
                |||||| ||||||||||||||
Db       1441 AGAACCTACAGTCTCAGTGA 1460


RESULT 4
US-08-836-620A-1
; Sequence 1, Application US/08836620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/03933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607532.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2568 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus rattus
US-08-836-620A-1

  Query Match             65.3%;  Score 1239.6;  DB 2;  Length 2568;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1454;  Conservative    0;  Mismatches  279;  Indels    7;  Gaps    4;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              || || ||| |||||||| ||    | | ||| |||||||  ||||| |||||||||
Db        213 CATGAGTCTCTGAGAACATAAT--GTCCATCTGTACCTCTTCTCACAAGGAGTTTTCTCA 270

Qy         61 GCTGTTA---TCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTC 117
               ||||   |   ||| |||||||| || |||||||||| || ||||||| ||| ||| |
Db        271 GCTGCGACCCTCTGAAGACATGGAGATCAAAAACTCACCGTCGAGCCTTAGTTC-CCTGC 329

Qy        118 CTCCTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTC 177
                |||||  |||||  || || |||||| ||||||||||||||||  |||| |||||  |||||
Db        330 TTCCTATAACTGTAGCCAGTCCATCCTACCCCTGGAGCACGGCCCCATCTACATCCCTTC 389

Qy        178 CTCCTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGAT 237
              |||||| |||||||  || ||||||  ||| ||||  |||||||||||  || ||||||||||
Db        390 CTCCTACGTAGACAACCGCCATGAGTATTCAGCTATGACATTCTACAGTCCTGCTGTGAT 449

Qy        238 GAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCAC 297
```

```
                ||| |||||  |||||  |||    || |||| |||| ||||||||||  || | || |||
Db        450 GAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGCAC 509

Qy        298 AAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCA 357
              |||||||||||| | ||||||||  |||||||||| |||||||||| |  |||| ||||
Db        510 AAGCCCAAAIGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGCCA 569

Qy        358 GTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGA 417
              | ||| |  || |||||  |||||||||||||||||||||||||||||||||||||||||||||
Db        570 ATCATCGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTAGA 629

Qy        418 ACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGC 477
              |||||||||||||||||||||||||||||||||||| |||||| |||||||| | |||| ||
Db        630 GCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGTGC 689

Qy        478 CAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGA 537
              |||||||||||||| |||||  ||||||||||||||||||||||| | ||||||||||||
Db        690 CAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGCGA 749

Qy        538 TTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAA 597
              ||| |||||  || |||||  || || || |||||  ||||||||||||||||||||||||||||
Db        750 TTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTTAA 809

Qy        598 AACAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGA 657
              |||||||||||||||||||||||||||||||| |||||||  || ||||||||||| || ||
Db        810 AAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATAGA 869

Qy        658 TAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAAT 717
              ||  |||||||| || ||||||||||||||||||||||||||| |||||||| |||| ||||
Db        870 CAACAACCCGCCTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTCTTATGAACTACCAAT 929

Qy        718 GCTCAAGTGTGGCTCCCGCAGCAGCAGCATCTGGGTACCGCCTTCTGCGCAGCACAGAAG 777
              ||| ||||||||  ||| |||||||  | ||||||||||  | |||||||| ||||||||
Db        930 GCTCAAGTCTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGGCAGAGAAG 989

Qy        778 TGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCG 837
              | |   ||||||| | ||||      |||| ||||||||||  || || || || |||||
Db        990 TTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCCCG 1049

Qy        838 AGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGA 897
              |||  |||||| ||||||  | || |||| || |||| || ||||||||||||||||||||
Db       1050 GGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTGGA 1109

Qy        898 GGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCAT 957
              ||||| || ||| |||||||| | ||||| |||||   ||||||||||||||||||||||
Db       1110 AGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCCAT 1169

Qy        958 GATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAA 1017
              |||||||||||| || ||| |||| |||||||  |||| |||||||  ||||||||||
Db       1170 GATGATGTCCCTCACTAAGCTGGCCGACAAGGAACTGGTGCACATGATTAGCTGGGCCAA 1229

Qy       1018 GAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTG 1077
              ||| || || ||||||||||||||||||||||||| ||||||| ||||||| || |||||
Db       1230 GAAAATGCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTAGAAAGCTC 1289

Qy       1078 TTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCT 1137
              |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||||
Db       1290 CTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAGCT 1349

Qy       1138 CATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCT 1197
              ||| || |||||  || || ||||||||||||||||||||||| |||||||||| |||||
Db       1350 CATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAACTGCGTAGAAGGGATTCT 1409

Qy       1198 GGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACA 1257
              |||||||||||||||||||||||||| || || ||||||| || ||||||||||||| ||
Db       1410 GGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAGCA 1469

Qy       1258 CAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGT 1317
              ||| || |||||||||||| |||||||||||||| ||||| ||||||||||||||  |||
Db       1470 CAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTGGC 1529

Qy       1318 CACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGAC 1377
                | || | |||||| || || || ||||||||||| | ||| | ||||||| |||||
```

```
Db        1530 TTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTGAC 1589

Qy        1378 CGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCG 1437
               ||||| |||| |||||||||||| ||||| || ||||||||||||||||| || | ||
Db        1590 AGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTCCG 1649

Qy        1438 CCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGA 1497
               ||||| |||||||||||||| || || |||||||||||     ||||||||||||||||
Db        1650 ACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCACATCAGTAACAAGGGCATGGA 1709

Qy        1498 ACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGAT 1557
               |||||||||||| ||||||||||||||||||||||| ||||||||||||||||||||||
Db        1710 ACATCTGCTCAGCATGAAGTGCAAAAATGTGGTCCCGGTGTATGACCTGCTGCTGGAGAT 1769

Qy        1558 GCTGAATGCCCACGTGCTTCGCCGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCC 1617
               ||||||||| ||| |||||| ||| ||||||||| ||| ||||||| ||||||||| |
Db        1770 GCTGAATGCTCACACGCTTCGAGGGTACAAGTCCTCAATCTCGGGGTCTGAGTGCAGCTC 1829

Qy        1618 GGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGC 1677
               |||||||||||||| | ||||||| |||||||||||| |||||||||||| | | |
Db        1830 AACAGAGCACAGTAAGAACAAAGAGAGCTCCCAGAACCTACAGTCTCAGTGATGGCCAGC 1889

Qy        1678 CCTGAGGTGAACTGGC-CCACAGAGGTCACAAGCTGAAGCGTGAACTCCAGTGTGTCAGC 1736
               ||||||| | || | |  || ||| |    ||  || | | || || | ||  | |  ||
Db        1890 CCTGAGGCGGACAGACTACAGAGATGGTCAAAAGTGGAACATGTACCCTAGCATCTGGGC 1949


RESULT 5
US-08-826-620A-6
; Sequence 6, Application US/08826620A
; Patent No. 5958710
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Orphan receptor
;    NUMBER OF SEQUENCES:  19
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/826,620A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/00933
;      FILING DATE:
;      APPLICATION NUMBER:  GB 9518272.1
;      FILING DATE:  08-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9605550.4
;      FILING DATE:  15-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9607542.0
;      FILING DATE:  11-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9609576.5
;      FILING DATE:  08-MAY-1996
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1458 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-826-620A-6

  Query Match             57.9%;  Score 1099.6;  DB 2;  Length 1458;
  Best Local Similarity   84.6%;  Pred. No. 1.2e-292;
  Matches 1234;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy         212 ATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCCCAGCAATGTCACTAACTTC 271
               ||| ||||||| || |||||||||||||| |||||  |||||||||    |  |||| ||

Db           1 ATGGCATTCTACAGTCCTGCTGTGATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTG 60

Qy         272 GAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCAC 331
               |||||||||||||| ||| |||||  |||||||||||||| | ||||||||  |||| |||
Db          61 GAAGGTGGGCCTGTTCGCCAGACTGCAAGCCCAAATGTGCTATGGCCAACTTCTGGACAC 120

Qy         332 CTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTATGCGGAACCTCAAAAGAGT 391
               || ||||||||||    ||| |||| | ||| | ||| ||||| ||||||||||||||||
Db         121 CTCTGTCCTTTAGCCACCCACTGCCAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGT 180

Qy         392 CCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGTAAACAGAGAGACACTGAAA 451
               || |||||||||||||||| |||||||||||| |||||||||||||||||||||| |||||
Db         181 CCTTGGTGTGAAGCAAGATCACTAGAACACACCTTGCCTGTAAACAGAGAGACCCTGAAG 240

Qy         452 AGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGAT 511
               ||||||  || ( |||| | |||| ||||||||||||||| ||||| |  | |||||||||
Db         241 AGGAAGCTTGGCGGGAGCGGTTGTGCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGAT 300

Qy         512 GCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCG 571
               |||||||||| || |||||||| |||| |||| || |||| || || ||||||||
Db         301 GCTCACTTCTGTGCCGTCTGCAGTGATTATGCATCTGGGTATCATTACGGTGTCTGGTCC 360

Qy         572 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGT 631
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db         361 TGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGACTATATCTGT 420

Qy         632 CCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGA 691
               ||||| || |||||||||| || || || ||||||| || | ||||||||||||||||||
Db         421 CCAGCCACGAATCAGTGTACGATAGACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGA 480

Qy         692 CTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGG 751
               ||||| |||||||||||||| |||||||| |||||||| ||| ||||||| || ||||||
Db         481 CTTCGCAAGTGTTACGAAGTAGGAATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGG 540

Qy         752 TACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCC 811
               |||||  | || || |||||||||||||||  ||||||| |||| ||      ||| |||
Db         541 TACCGAATAGTACGAAGACAGAGAAGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCC 600

Qy         812 AAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCC 871
               |||||||   | || ||| | |||| |||  |||||| |||||| || | |||| |||
Db         601 AAGAGAACCAGTGGGCACACACCCCGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCC 660

Qy         872 GAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCC 931
               |||||||| |||||||||| |||| |||||||| ||| |||||||  | || || |||
Db         661 GAGCAGCTGGTGCTCACCCTGCTGGAAGCTGAGCCACCCAATGTCCTACTGAGTCCTCCC 720

Qy         932 AGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAG 991
               ||   |||||||||||||||||||||||||||||||||||| || ||| |||| ||||||||
Db         721 AGCATGCCCTTCACCGAGGCCTCCATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAA 780

Qy         992 TTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTC 1051
               |||| ||||||||  |||||||||||||| || || ||||||||||||||||||||||||
Db         781 CTGGTGCACATGATTGGCTGGGCCAAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTG 840

Qy        1052 GACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGG 1111
               |||||||| || |||||||| |||| |||||||||||| | ||| ||||||||||||||
Db         841 GACCAAGTCCGCCTCTTGGAAAGCTGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGG 900

Qy        1112 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGAT 1171
               ||||| || |||||||||||||||||||||||||||||||||||| || ||||||||||||
Db         901 CGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGAT 960

Qy        1172 GAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCA 1231
               ||||||| ||||| |||| |||||||||||||||||||||||||||| || ||  ||
Db         961 GAGGGGAAGTGCGTGGAAGGGATTCTGGAAATCTTTGACATGCTCCTGGCGACGACGGCA 1020

Qy        1232 AGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTG 1291
               |||| || |||||||||| || |||||||||||||| |||||  |||||||||| ||
Db        1021 CGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTCAAGGCCATGATTCTC 1080

Qy        1292 CTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAG 1351
               ||||| ||||||||||||   ||| || || | |||||| || || || ||||||||
Db        1081 CTCAACTCCAGTATGTACCACTTGGCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAG 1140
```

```
Qy       1352 CTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAACAGCCGC 1411
             ||| | ||| |  ||||||| ||||| ||||  |||| |||||||| | |||||  |
Db       1141 CTGACACACCIATTGAACGCAGTGACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGA 1200

Qy       1412 ATCTCCTCCCAGCAGCAATCCATCGCCCTGCCTAACCTCCTGATCGTCCTGTCCCACGTC 1471
             ||||| |||||||||||| ||  | || |||| |||||||||||||| || || || |||
Db       1201 ATCTCTTCCCAGCAGCAGTCAGTCCGTCTGGCCAACCTCCTGATGCTTCTTCTCATGTC 1260

Qy       1472 AGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTC 1531
             |||||    ||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1261 AGGCACATCAGTAACAAGGGCATGGAACATCTGCTCAGCATGAAGTGCAAAAATGTGGTC 1320

Qy       1532 CCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCC 1591
             || ||||| |||||||||||||||||||||||||||| ||| |||||| |||| |||||||
Db       1321 CCGGTCTACGACCTGCTGCTGGACATGCTGAATGCTCACATGCTTCGAGGTACAAGTCC 1380

Qy       1592 TCCATCAGGGGTCCGAGTGCAGCCCCCCACAGCAGCAGTAAAAGCAAACAGGGTCCCAG 1651
             || ||| ||||||| | |||| || || ||||||||||||| ||||||||||||||||||
Db       1381 TCAATCTCGGGGTCTGGGTGCTGCTCCACAGAGGACAGTAAGACAAACAGGGTCCCAG 1440

Qy       1652 AACCCACAGTCTCAGTGA 1669
              ||||  ||||||||||||
Db       1441 AACCTCCAGTCTCAGTGA 1458


RESULT 6
US-08-076-726-16
; Sequence 16, Application US/08076726
; Patent No. 5464758
;  GENERAL INFORMATION:
;    APPLICANT:  Gossen, Manfred
;    APPLICANT:  Bujard, Hermann
;    TITLE OF INVENTION:  Tight Control of Gene Expression in
;    TITLE OF INVENTION:  Eucaryotic Cells by Tetracycline-responsive Promoters
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  District of Columbia
;      COUNTRY:  United States of America
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,726
;      FILING DATE:  14-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942 2490001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;      TELEX:  248636 SSK
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4963 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
US-08-076-726-16

  Query Match             20.6%;  Score 490.8;  DB 1;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 1.7e-97;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACIATGGAGIC 565
```

```
Db       1003 AAGCAGACTCGCTACTGTGCCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGCAGTC 1062

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| |||| || || |||||||| || || |||| ||||||||||||| || |||
Db       1063 TGCTGCTGTGACGCGTGCAAGCCCTTCTTCAAGAGAAGTATTCAAGGACATAACCACTAT 1122

Qy        626 ATTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || ||||||||| || | ||||||||||||||
Db       1123 ATGTCTCCAGCCACCAACCAGTGCACCATTGATAAAAACACCAGCAACAGCCTGCCAGCC 1182

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db       1183 TGCCCGCTCCGCAAATCCTACGAAGTGGGAATGATCAAAGCTGGGATACGAAAAGACCGA 1242

Qy        746 TGTCC------------GTACCGCCTTGTGCCGGAGACACACAAGTCCCCACAGACAGCTC 793
              | ||              | | | |      | |||| |    |  | | | |
Db       1243 AGACGACGGACAATGTTGAAACACAAGCCCCAGAGACATGATGGGGAGGGPAGGGTCAA 1302

Qy        794 CACTGTGCCCGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 843
                | | |   | || |||| || |  ||   |||  |   ||
Db       1303 GTGGGGTCTGCTGCAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        844 ----------GCAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                       | |   |||       ||||||   |||| ||| | ||  |  || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGCGCTTG 1422

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||   |     |           || |     |||||||  ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  |||||||| |||| |||| |||| |||| |||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||   || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |   ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  ||||  | |||||||| |  |  |  |||||||| |  |||| |||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTCGAGAGTTAAAA 1249
              || ||  |||| |||||  |||||||| |||| ||  | ||  |||| ||   | || ||
Db       1723 GGCATGCTGGAGATCTTTGGACATGCTGCTGGCCACATCATCTCGGTTCGGCATCATCAAT 1782

Qy       1250 CTGCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||    | || | | | || |||| | || ||  |||| |||||  |  |||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGGCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  |  |    |||| | |  ||   ||    |     |  |   |  | |||
Db       1843 ACATTTCTGTCCAGCAGCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| ||||||  |||||||    |||||||
Db       1963 CAGCACCAGCGCCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| |||||| || ||||||||||||| || |||| ||  | |||||||||
Db       2023 AAAGCCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||  | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGGCCCACAGGCTGCACGCCGGCAGCTGCCCC 2122
```

```
Db        2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 7
US-08-260-452-9
; Sequence 9, Application US/08260452
; Patent No. 5650298
;  GENERAL INFORMATION:
;     APPLICANT:  Gossen, Manfred
;     APPLICANT:  Bujard, Hermann
;     APPLICANT:  Salfeld, Jochen
;     APPLICANT:  Voss, Jeffrey
;     TITLE OF INVENTION:  Tight Control of Gene Expression in Eucaryotic
;     TITLE OF INVENTION:  Cells by Tetracycline-responsive Promoters
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lahive & Cockfield
;       STREET:  60 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/260,452
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/076,427
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giulio A. DeConti, Jr.
;       REGISTRATION NUMBER:  31,503
;       REFERENCE/DOCKET NUMBER:  BBI-013CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4963 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR4
US-08-260-452-9

  Query Match             20.6%;  Score 490.8;  DB 1;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 1.7e-97;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy         506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
               | |||  ||| || ||| || || ||||  || || || || || || || ||||||||||
Db        1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy         566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
               ||||| ||||| || |: |||||||| || || ||||| |||||||||||||| || |||
Db        1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy         626 ATTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
               || |||||||| || || ||||| || || |||||||||| || | ||||||||||||||||
Db        1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy         686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
               ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db        1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy         746 TGTGG------      GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
               | ||           | | | |     | |||| |     | |  | | |
Db        1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy         794 CACTGTGCCGGCCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG- ------ 843
               | | |     | || |||| || | ||   ||| |  ||
Db        1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy         844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                     | |   |||      |||||||  |||| ||| | ||  |  || |
Db        1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCATGGCCTTG 1422

Qy         893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
               |||| |||||||| ||| |  |  |        || |    ||||||| ||
Db        1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy         950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
               || || ||||||  | ||||||||  |||| |||| |||| |||| |||||||||| |
Db        1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy        1010 TGGGCCAAGAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
               ||||| |||| | | || ||||||||||||  | | |||   || || || |  || |
Db        1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGCATTCGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy        1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
               ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db        1603 GAATGTGCCTGGCTAGATCTCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy        1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
               |  ||||  | |||||||| |  |  |  |||||||| |  |||| |||| ||||
Db        1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy        1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTCGAGAGTTAAAA 1249
               || ||  |||| |||| ||||||| |||| ||  |  ||  |||| ||   | | ||
Db        1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy        1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
               || ||     | || | | | ||  |||| | || ||  |||| |||| |  ||||
Db        1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy        1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTCCTG 1366
               |  |  |  |   |||| | |  ||   ||  |  |  |  |  |  |  | |||
Db        1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy        1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
               ||    | || ||  |||| |    |||| ||||||   ||| |  ||   ||||||
Db        1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy        1427 CAATCCATGCGCCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
               ||   |  ||| ||||| | |||||  |  |||| |||||| |||||||   |||||||
Db        1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy        1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
               || |||||||| |||||  || ||||||||||||| || ||||| ||  | |||||||||
Db        2023 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy        1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCA 1586
               |||||||||||||| | ||||||   || |  | |  ||
Db        2083 CTGCTGGACATCCTCGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 8
US-08-481-970-9
; Sequence 9, Application US/08481970
; Patent No. 5859310
;  GENERAL INFORMATION:
;     APPLICANT:  Gossen, Manfred
;     APPLICANT:  Bujard, Hermann
;     APPLICANT:  Salfeld, Jochen
;     APPLICANT:  Voss, Jeffrey
;     TITLE OF INVENTION:  Animal Transgenic for a Tetracycline Controlled    Transcri
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Lahive & Cockfield
;       STREET:  60 State Street, Suite 510
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/481,970
;       FILING DATE:
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/360,452
;       FILING DATE:  14-JUN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/076,327
;       FILING DATE:  14-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giulio A. DeConti, Jr.
;       REGISTRATION NUMBER:  31,503
;       REFERENCE/DOCKET NUMBER:  BBI-013CP2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4963 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR4
US-08-481-970-9

  Query Match             20.6%;  Score 390.8;  DB 2;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 1.7e-97;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || ||||||||||
Db       1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGCAATGGTGAAGTGTGGCTCCCGGAGACAGACA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGCAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        746 TGTGG-----------GTACCGCCTTGTGCGGAGACAGACAAGTGCCGACGACGACGTG 793
               | ||          | | | |      | |||| |     | |  | |  |
Db       1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy        794 CACTGTGCCGCCAAGGCCAAGACAAGTGCGCGCCACCCCCCCCCAGTGCG----- ---- 843
                | | |     | || ||||  || | |  ||    |||  |   ||
Db       1303 GTGGGGTCTGCTGGAGACATCAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                       | |   |||       ||||||  |||| ||| | ||  |  || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGCGCCTTG 1422

Qy        893 CTGCAGCCTGAGCCGCCCCATGTGCTGATC---ACCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||    |     |        || |     |||||||  ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||    | ||||||||  |||| |||| |||| |||| ||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy       1070 GAGACCTGTTGGATGGAGGTGTTAATGATGGGGCTGATCTGGCGCTCAATTGACCACCCC 1129
              || |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGACATCCTGATGATTGGTCTGGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  ||||| | |||||||  |  |  |  |||||||| |  |||| ||||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTGGAGAGTTAAAA 1249
              || ||  |||| ||||| |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTTGACATGCTGCTGGCTACATCATCTGGCTTCGGCATGATGAAT 1782

Qy       1250 CTGCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||    | || | | | | ||  |||| | || ||  |||| |||||  |  |||||
Db       1783 CTGCAGCGAGAGCAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATCTGGAGTCTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              | | | | |  |||| | |  ||  ||  |    |  |  |  | | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  ||||| |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGGCGCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| ||||| | ||||| | |||| |||||| |||||||   |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| |||||| || ||||||||||| || ||||| ||  | |||||||||
Db       2023 AAAGCCATCGAGCATCTGTACACCATGAAGTGCAACAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGCAGATGCTGGACGCCCACCGCCTACATCCCCCCA 2122


RESULT 9
US-08-897-719-9
; Sequence 9, Application US/08897719
; Patent No. 5922927
;  GENERAL INFORMATION:
;    APPLICANT:  Gossen, Manfred
;    APPLICANT:  Bujard, Hermann
;    APPLICANT:  Salfeld, Jochen
;    APPLICANT:  Voss, Jeffrey
;    TITLE OF INVENTION:  Tight Control of Gene Expression in Eucaryotic
;    TITLE OF INVENTION:  Cells by Tetracycline-responsive Promoters
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive & Cockfield
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
```

```
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/897,719
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/260,452
;       FILING DATE:
;       APPLICATION NUMBER:  08/076,327
;       FILING DATE:  14 JUN 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giulio A. DeConti, Jr.
;       REGISTRATION NUMBER:  31,503
;       REFERENCE/DOCKET NUMBER:  BB1-013CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227 5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4963 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR4
US-08-897-719-9

  Query Match             20.6%;  Score 399.8;  DB 2;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 1.7e-97;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || |||||  || || || || || || || |||||||||
Db       1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || ||||| || || ||||||||| || | ||||||||||||||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        746 TGTGG----------GTACCGCCTTGTGCGGACACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||          | | | |     | |||| |     | |  | |  |
Db       1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGACAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG------ 843
              |  | |     | || ||||  || |  ||    |||  |   ||
Db       1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        844 -------GGAGCTGCTGCTGGACGCCCTGACCCCCGAGCAGCTAGTGCTCACCCTC 892
                    | |    |||       ||||||   |||| ||| | ||  |  || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCATGGCCTTG 1422

Qy        893 CTGCAGGCTGAGCCGCCCCATGTGCTGATC   ACCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| ||||||| |||    |      |        || |      ||||||| ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   | ||||||||  |||| |||| |||| |||| |||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1602

Qy       1070 GACAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||  | ||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  |||||  | ||||||||  |  |  |  |||||||| |  |||| ||||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| ||||| |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||      | || | | | ||  |||  | || ||  |||| ||||  |  |||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
                | | | |    |||| | |  ||   ||    |       | |    |  | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
                ||    | || ||  ||||| |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || ||||||||||||| || ||||| ||  | |||||||||
Db       2023 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGTCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              ||||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 10
US-09-163-269-9
; Sequence 9, Application US/09163269
; Patent No. 6252136
;  GENERAL INFORMATION:
;    APPLICANT:  Gossen, Manfred
;    APPLICANT:  Bujard, Hermann
;    APPLICANT:  Salfeld, Jochen
;    APPLICANT:  Voss, Jeffrey
;    TITLE OF INVENTION:  Animal Transgenic for a Tetracycline-Controlled    Transcri
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive & Cockfield
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/163,269
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/481,970
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076,327
;      FILING DATE:  14-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Giulio A. DeConti, Jr.
;      REGISTRATION NUMBER:  31,503
```

```
;       REFERENCE/DOCKET NUMBER:  BRI-013CP2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4963 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR4
US-09-163-269-9

  Query Match             20.6%;  Score 390.8;  DB 4;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 3.7e-97;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || || |||||||||
Db       1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGGCTTCTTCAAGAGAAGTATTCAAGGACATAACCACTAT 1122

Qy        626 ATTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || ||||| || || |||||||||| || | ||||||||||||||||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTCATAAAAACAGGAGCAAGAGCTGCCAGGCC 1182

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || |||||||||||||||| ||||  ||||    || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGGTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        746 TGTGG----------GTACCGCCTTGTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
               | ||          | | | |     | |||| |     | |  | | |
Db       1243 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCGCCCCCGAGTGCG---------- 843
                | | |     | || |||| || |  ||   |||  |   ||
Db       1303 GTGGGCTCTGCTGGAGACATGACAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        844 ----------GGAGCTGCTGCTGGAGCCCTGAGCCCGAGACACTAGTGCTCACCCTC 892
                       | |   |||      ||||||  |||| ||| | | ||  |  || |
Db       1363 CGCTCTAAGAACAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCATGGCCTTG 1422

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| ||||||||| ||| |   |      |          || |     ||||||| ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  |||||||  |||| ||| |||| |||| ||||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| ||| | | || ||||||||||||  | | |||    || || | |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCACCTTCTA 1602

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATCGTATGAA 1189
              |  ||||  | |||||||  |  |  |  ||||||| |  |||| |||| ||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTCTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db       1723 GCCATCCTCGAGATCTTCGACAAGCTCTCGGTGACATCATCTCGGTCGGCATCATGAA- 1781

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||| | || ||  |||| |||||  |  |||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  | |  |||| | | ||  ||   |  |  |  |    | | |  | | |||
Db       1843 ACATTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||    |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| |||||| || ||||||||||||| || ||||| ||  | |||||||||
Db       2023 AAAGGCATCCACCACATCTCTACAGCATCAACTCCAACAACGTGGTCGTCTCTATGACCTC 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTTGGAGGCCCACGGCCTACATGCGCCCA 2122


RESULT 11
US-09-041-886-34
; Sequence 34, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6450 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  361..2146
US-09-041-886-34

  Query Match             20.5%;  Score 389.2;  DB 4;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 5.3e-97;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;
```

```
Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db       1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || |||||||||||||||| |||| |||| || | ||| ||
Db       1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1137

Qy        746 TGTGG---------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||           | | | |      | |||| |    | |  | |  |
Db       1138 AGAGGAGGCAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGTGAA 1197

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG-------- 843
              | | |   | || |||| || |  ||   ||| |  ||
Db       1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTGGCCAAGCCCGCTCATGATCAAA 1257

Qy        844 --      GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                  | |   |||      ||||||  |||| ||| | ||     || |
Db       1258 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCTTG 1317

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||   |     |          || |     |||||||  ||
Db       1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  ||||||||  |||| |||| |||| |||| ||||||||| |
Db       1378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || |||||||||||  | | |||    || || || |  ||  |
Db       1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |   ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1498 GAATGTGCCTGGCTAGACATCCTGATGATTGGTCTCGTCTGGCGCTCCATGCAGCACCCA 1557

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAA 1189
              |  |||||  | ||||||||  |  |  |  |||||||| |  |||| |||||  |||||
Db       1558 GTGAAGCTACTGTTTGTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| ||||| |||||||| ||||| ||  |  || |||| ||  |  || ||
Db       1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677

Qy       1250 CTCCAACACAAACAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  ||||  | || ||  |||| |||||  |  ||||||
Db       1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGATGTAC 1737

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACACCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  |  |   |||| | |  ||   ||    |  |   |   | | |||
Db       1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy       1367 AACGCCGTGACCGATGCTTTGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  ||||| |     |||| ||||||   ||| |  ||   ||||||
Db       1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| ||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || ||||||||||||| || |||||  ||  | |||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||| | ||||||   || |  | |  ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 12
US-08-453-998-1
; Sequence 1, Application US/08453998
; Patent No. 6444438
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, PIERRE
;    APPLICANT:  METZGER, DANIEL
;    APPLICANT:  WHITE, JOHN
;    TITLE OF INVENTION:  METHOD FOR THE PREPARATION OF A PROTEIN
;    TITLE OF INVENTION:  BY YEASTS USING AN INDUCIBLE SYSTEM, VECTORS AND
;    TITLE OF INVENTION:  CORRESPONDING TRANSFORMED STRAINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVE., N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,998
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/161,064
;      FILING DATE:  03-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CHAPIN, MARIANA K.
;      REGISTRATION NUMBER:  35,843
;      REFERENCE/DOCKET NUMBER:  1037/98493
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861 3711
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6450 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-453-998-1

  Query Match             20.5%;  Score 389.2;  DB 4;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 5.4e-97;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db       1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || |||||||||||||||| |||| |||| || | ||| ||
Db       1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1137
```

```
Qy        746 TGTGG------------GTACCGCCTTGTGCCGAGACAGAGAAGTGCCGACGAGCAGCTG 794
              | ||              | | | |       | |||| |     | | | | | |
Db       1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGCTGAA 1197

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 844
              |  | |     | || |||| || |  ||    ||| |   ||
Db       1198 GTGGGGTCTGCTGGAGACATGAGACGTGCCAACCTTTGGCCAAGCCCGCTCAIGATCAAA 1257

Qy        844 ----------GGAGCTGCTGCTGCAGCGCCTGAGCCCGGAGCAGCTAGTGCTCACCCTC 892
                       | |    |||        |||||   |||| ||| | ||     || |
Db       1258 GGCTCTAAGAAGAACAGCCTGGCCTTGTCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1317

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||    |       |          || |     |||||||  ||
Db       1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  |||||||  |||| |||| |||| |||| ||||||||| |
Db       1378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGAGCTGGTTCACATGATCAAC 1437

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||   || || || |  ||  |
Db       1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1498 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1557

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATCAGGGGAAATGCGTAGAA 1189
              |  ||||  | |||||||  |  | |  ||||||| | ||||| ||||| |||||
Db       1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTCGAGAGTTAAAA 1249
              || ||  |||| ||||| |||||||| ||||| ||  | ||  ||| ||   | | ||
Db       1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  ||||  | || ||  |||| ||||  |  |||||
Db       1678 CTGCAGGAGAGGAGTTTGTGTGCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1737

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  |  |    |||| | |  ||   ||    |      | |     |  | |||
Db       1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1798 GACAAGATCACAGACACTTTGATCCAGCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATCGAGTAAC 1486
              ||   |  ||| |||||  | ||||  |  |||| |||||| |||||||   |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || ||||||| ||||||  || ||||||||||| || |||| ||  | ||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||| | ||||||   || | | ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 13
US-08-693-940-2
; Sequence 2, Application US/08693940
; Patent No. 6133027
;  GENERAL INFORMATION:
;    APPLICANT:  Yee, Jiing-Kuan
;    APPLICANT:  Friedman, Theodore
;    APPLICANT:  Chen, Shin-Tai
;    TITLE OF INVENTION:  Inducible Expression System
;    TITLE OF INVENTION:  Useful in the Generation of Packaging Cell Lines for
;    TITLE OF INVENTION:  Pseudotyped Retroviral Vectors
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/693,940
;      FILING DATE:  07-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Francis, Carol L
;      REGISTRATION NUMBER:  36,513
;      REFERENCE/DOCKET NUMBER:  6510-055001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3400
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1956 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...1953
;      OTHER INFORMATION:
US-08-693-940-2

  Query Match             15.0%;  Score 285.6;  DB 3;  Length 1956;
  Best Local Similarity   63.8%;  Pred. No. 9.1e-69;
  Matches  467;  Conservative    0;  Mismatches  259;  Indels    6;  Gaps    2;

Qy        861 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 920
              ||||||   |||| ||| | ||     || |  |||| |||||||| |||    |
Db       1094 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1153

Qy        921 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 977
              |    |  || |     |||||||  || || || ||||||   |  ||||||||
Db       1154 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1213

Qy        978 TGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 1037
              |||| |||| ||| |||| |||||||||| |||||| |||| | | || ||||||||||
Db       1214 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1273

Qy       1038 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 1097
              |  | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1274 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1333

Qy       1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
              | || ||  | |||||||| || || ||||| |  |||||  | ||||||||  | |
Db       1334 TTGGTCTCCTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1393

Qy       1158 TTCTGGACAGGGATGAGGGAAATGTGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
              |  |||||||| |  |||| ||||| ||||| || ||  |||| |||| |||||||| |
Db       1394 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGCCATGGTGGAGATCTTCGACATGCTGC 1453

Qy       1218 TGGCAACTACTTCAAGGTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1277
              |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||  |||
Db       1454 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCTCA 1513

Qy       1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1336
              |  | || ||  |||| ||||| |  ||||| |  |  |  |    |||| | |  ||
```

```
Db       1514 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1573

Qy       1337   GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1394
                ||    |     |   |     |   | ||| ||    | || ||  ||||| |
Db       1574 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1633

Qy       1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
              |||| ||||||    ||| |  ||   ||||||||   |  ||| |||||  | |||||
Db       1634 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1693

Qy       1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1514
              |  |||| |||||| |||||||   |||||||||| |||||||| ||||||  || |||||
Db       1694 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1753

Qy       1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
              ||||||| || |||| ||  | ||||||||||||||||||||||| | ||||||   |
Db       1754 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCC 1813

Qy       1575 TTCGCGGGTGCA 1586
              | |  | |  ||
Db       1814 TACATGCGCCCA 1825


RESULT 14
US 09 566 660 2
; Sequence 2, Application US/09566660
; Patent No. 6432705
; GENERAL INFORMATION:
;  APPLICANT:  Yee, Jiing Kuan
;  APPLICANT:  Friedmann, Theodore
;  APPLICANT:  Chen, Shin-Tai
;  TITLE OF INVENTION:  Inducible Expression System
;  FILE REFERENCE:  6510-055CON
;  CURRENT APPLICATION NUMBER: US/09/566,660
;  CURRENT FILING DATE:  2000 05 08
;  PRIOR APPLICATION NUMBER:  08/693,940
;  PRIOR FILING DATE:  1996-08 07
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1956
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1953)
;   OTHER INFORMATION: Multi chimeric transactivating factor
US 09 566 660 2

  Query Match             15.0%;  Score 285.6;  DB 4;  Length 1956;
  Best Local Similarity   63.8%;  Pred. No. 9.1e-69;
  Matches  467;  Conservative    0;  Mismatches  259;  Indels    6;  Gaps    2;

Qy        861 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 920
              ||||||   |||| ||| | ||      || |  |||| |||||||| |||    |
Db       1094 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1153

Qy        921 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 977
              |      || |     |||||||  || || || ||||||   |  ||||||||
Db       1154 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1213

Qy        978 TGGCCGACAAGGAGTTGGTGCACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 1037
              |||| |||| |||| |||| |||||||||| |||||| ||| | | || ||||||||||
Db       1214 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1273

Qy       1038 AGCTCAGCCTGTTCGACCAACTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 1097
              |  | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1274 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1333

Qy       1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
              | || ||  | |||||||| || || ||||| |  |||||  | ||||||||  |  |
Db       1334 TTGGTCTCGGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1393

Qy       1158 TTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
              |  |||||||| |  |||| ||||| ||||| || ||  |||| ||||| |||||||| |
Db       1394 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1453

Qy       1218 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTTCAACACAAAGAATATCTCTGTGTCA 1277
              |||| ||  | ||  |||| ||   | | || || ||      | || | | | ||  |||
Db       1454 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1513

Qy       1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT  1336
              |  | ||  ||  |||| |||||  |  |||||  |  |  |  |  |    |||| | |  ||
Db       1514 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1573

Qy       1337   GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1394
                ||    |     |   |     |   | ||| ||    | || ||  ||||| |
Db       1574 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1633

Qy       1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
              |||| ||||||    ||| |  ||   ||||||||   |  ||| |||||  | |||||
Db       1634 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1693

Qy       1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1514
              |  |||| |||||| |||||||   |||||||||| |||||||| ||||||  || |||||
Db       1694 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1753

Qy       1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
              ||||||| || |||| ||  | ||||||||||||||||||||||| | ||||||   |
Db       1754 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCC 1813

Qy       1575 TTCGCGGGTGCA 1586
              | |  | |  ||
Db       1814 TACATGCGCCCA 1825


RESULT 15
US 08-564-264-2
; Sequence 2, Application: US/08564264
; Patent No. 6040430
;  GENERAL INFORMATION:
;    APPLICANT:  STEWART, Francis
;    TITLE OF INVENTION:  REGULATION OF SITE-SPECIFIC
;    TITLE OF INVENTION:  RECOMBINATION BY SITE-SPECIFIC RECOMBINASE/NUCLEAR
;    TITLE OF INVENTION:  RECEPTOR FUSION PROTEINS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/564,264
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02088
;      FILING DATE:  28 JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 93 110 298.2
;      FILING DATE:  28 JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murray, Robert B.
;      REGISTRATION NUMBER:  22,980
;      REFERENCE/DOCKET NUMBER:  P564-5019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)638-5000
;      TELEFAX:  (202)638-4810
;  INFORMATION FOR SEQ ID NO:  2:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2322 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (1-1269)
;       OTHER INFORMATION:  /note= "FLP recombinase domain."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (1270-1284)
;       OTHER INFORMATION:  /note= "Linker peptide."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (1285-2322)
;       OTHER INFORMATION:  /note= "Estrogen binding domain."
US-08-564-264-2

  Query Match             15.0%;  Score 285.6;  DB 3;  Length 2322;
  Best Local Similarity   63.8%;  Pred. No. 9.9e-69;
  Matches  467;  Conservative    0;  Mismatches  259;  Indels    6;  Gaps    2;

Qy        861 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 920
              ||||||   |||| ||| | ||     || |  |||| |||||||| |||    |
Db       1460 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1519

Qy        921 TC---AGCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 977
              |         || |     |||||||  || || || ||||||    |  ||||||||
Db       1520 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1579

Qy        978 TGGCCGACAAGGAGTTGGTGCACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGC 1037
              |||| |||| |||| |||| |||||||||| |||||| ||| | | || ||||||||||
Db       1580 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1639

Qy       1038 AGCTCAGCCTGTTGCACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 1097
              | | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1640 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1699

Qy       1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
              | || ||  | |||||||| || || |||| |  ||||  | ||||||||  |  |
Db       1700 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1759

Qy       1158 TTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
              |  |||||||| |  |||| |||| |||| || ||  |||| |||| |||||||| |
Db       1760 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1819

Qy       1218 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTCTGTCA 1277
              |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||  |||
Db       1820 TGGCTACATCATCTCGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1879

Qy       1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCTCTCCTCACAGCGACCCAGGATCCT- 1336
              | | || ||  |||| |||||  |  ||||| |  |  |  |  |    |||| | |  ||
Db       1880 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1939

Qy       1337 --GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1394
                ||    |      | |    |   | ||| ||    | || ||  ||||| |
Db       1940 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1999

Qy       1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
              |||| ||||||   ||| |  ||   ||||||||   |  ||| ||||  | ||||
Db       2000 TGATGGCCAAGGCAGGCCTGACCTGCAGCAGCAGCAGCCAGCGCCTGGCCCAGCTCCTCC 2059

Qy       1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATCA 1514
              |  |||| |||||| |||||||   |||||||| |||||||| ||||||  || ||||
Db       2060 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 2119

Qy       1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
              |||||||  || |||||  ||  | ||||||||||||||||||||||| | ||||||   |
Db       2120 AGTGCAAGAACGTGGTGCCCCTCTATGACCTCCTGCTGGAGATGCTGGACGCCCACCGCC 2179

Qy       1575 TTCGCGGGTGCA 1586
              | |  | |  ||
```

```
Db       2180 TACATGCGCCCA 2191


RESULT 16
US-09-040-508-1
; Sequence 1, Application US/09040508
; Patent No. 6069239
;  GENERAL INFORMATION:
;    APPLICANT:  Mathias, Stephen L.
;    TITLE OF INVENTION:  NOVEL COMPOUNDS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,508
;      FILING DATE:  17-MAR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9705451.4
;      FILING DATE:  17-MAR-1997
;      APPLICATION NUMBER:  GB 9803289.9
;      FILING DATE:  16-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-30008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-040-508-1

  Query Match             6.0%;  Score 113.8;  DB 3;  Length 1615;
  Best Local Similarity   49.3%;  Pred. No. 1.4e-21;
  Matches  381;  Conservative    0;  Mismatches  362;  Indels   30;  Gaps    2;

Qy        527 GTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAG 586
              || ||  | ||   ||| || || || |||||||| ||    || |||||||  || |||
Db        545 GTGTGTGGTGACATCGCTTCTGGGTACCACTATGGGGTAGCATCATGTTGAAGCCTGCAAG 604

Qy        587 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 646
              || || || || || | |  |||||||  |||  || || |  || || || || ||| |
Db        605 GCATTCTTCAAGAGGACAATTCAAGGCAATATAGAATACAGCTGCCTGCCAGGAATGAA 664

Qy        647 TGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCGACTTCGGAAGTGTTAC 706
              |||  ||||   ||   | | || ||   ||||||||| |||||  |   ||||||||
Db        665 TGTGAAATCACAAAGCGCAGACGTAAATCCTGCCAGGCTTGCCGCTTCATGAAGTGTTTA 724

Qy        707 GAAGTGGGAATGGTCAAGTGTGGCTCCCGAGAGAGAGATGTGGGTACCGCCTTGTGCGG 766
              |||||| ||| ||||    ||    ||    || |||                    |
Db        725 AAAGTGGGCATGCTGAAAGAAGGGGTGCGTCTTGACAGA-   -----  ----------- 763

Qy        767 ACACACAGCAACTCCCCACGACCAGCTGCACTGTCCCGCCAACCCCAACACAACTCCCCGT 826
                  |||  |  ||||  | ||    | | |  |  |  |  |  ||
Db        764 ------GTACGTGGAGGTCGGCAGAAGTACAAGCGCAGGATAGATGCGGAGAACAGCCCA 817
```

```
Qy        827 CACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTC 886
              ||  |  ||    || ||  | ||       |    |       | ||    |   |
Db        818 TACCTGAACCCTCAGCTGGTTCAGCCAGCCAAAAAGCCATATAACAAGATTGTCTCACAT 877

Qy        887 ACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACC 946
                |  |||  |    |  |             ||       ||| | ||  |||  ||
Db        878 TTGTTGGTGGCTGAACCGGAGAAGATCTATGCCATGCCTGACCCTACTGTCCCCGACAGT 937

Qy        947 GAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATC 1006
              ||  |       | |  |||     | |||||||   |||||||    || ||
Db        938 GACATCAAAGCCCTCACTACACTGTGTGACTTGGCCGACCGAGAGTTGGTGGTTATCATT 997

Qy       1007 AGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTC 1066
               | ||||| ||| | ||||| |||||      |||   ||||   |||||  || | ||
Db        998 GGATGGGCGAAGCATATTCCAGGCTTCTCCACGCTGTCCCTGGCGGACCAGATGAGCCTT 1057

Qy       1067 TTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCAC 1126
               || ||||   |||||||||  | || ||  | ||  |  | |  || ||  ||
Db       1058 CTGCAGAGTGCTTGGATGGAAATTTTGATCCTTGGTGTCGTATACCGGTCTCTTTCATTT 1117

Qy       1127 CCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTA 1186
                |  | ||  ||| |||    ||| | |  |||||  ||  ||  ||||   ||
Db       1118 GAGGATGAACTTGTCTATGCAGACGATTATATAATGGACGAAGACCAGTCCAAAT--TA 1174

Qy       1187 GAAGGAATTCTGGAAATCTTTGACATGCTCCTGCCAACTACTTCAAGGTTTCGAGAGTTA 1246
                | |||  |||| ||  |   | |     |||||                | |      |
Db       1175 GCAGGCCTTCTTGATCTAAATAATGCTATCCTGCAGCTGGTAAAGAAATACAAGAGCATG 1234

Qy       1247 AAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTC 1299
              || ||  || |  |||||| | ||    |||| || ||    ||   |||||
Db       1235 AAGCTGGAAAAAGAAGAATTTGTCACCCTCAAAGCTATAGCTCTTGCTAATTC 1287


RESULT 17
US-09-500-654-1
; Sequence 1, Application US/09500654
; Patent No. 6359116
; GENERAL INFORMATION:
;  APPLICANT: Stephen L. Mathias
;  TITLE OF INVENTION: No. 6359116e1 Compounds
;  FILE REFERENCE: GP-30008 D1
;  CURRENT APPLICATION NUMBER: US/09/500,654
;  CURRENT FILING DATE: 2000-02-09
;  EARLIER APPLICATION NUMBER: 9705451.4
;  EARLIER FILING DATE: 1997-03-17
;  EARLIER APPLICATION NUMBER: 9803289.9
;  EARLIER FILING DATE: 1998-02-16
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1615
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-500-654-1

  Query Match             6.0%;  Score 113.8;  DB 4;  Length 1615;
  Best Local Similarity   49.8%;  Pred. No. 1.4e-21;
  Matches  481;  Conservative    0;  Mismatches  362;  Indels   40;  Gaps    2;

Qy        527 GTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAG 586
              || ||  | ||   ||| || || || || |||||||| ||    || |||||||  || |||
Db        545 GTGTGTGGTGACATCGCTTCTGGGTACCACTATGGGGTAGCATCATGTGAAGCCTGCAAG 604

Qy        587 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 646
              || || || || || |  |||||||| |||  || || |  || || || || ||| |
Db        605 GCATTCTTCAAGAGGACAATTCAAGGCAATATAGAATACAGCTGCCCTGCCACGAATGAA 664

Qy        647 TGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTAC 706
              |||  ||||   ||   | | || ||   |||||||| |||||  |  |||||||
Db        665 TGTGAAATCACAAAGCGCAGACGTAAATCCTGCCAGGCTTGCCGCTTCATGAAGTGTTTA 724

Qy        707 GAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGG 766
              |||||||  ||| ||||    ||    ||    || |||

Db        725 AAAGTGGGCATGCTGAAAGAAGGGTGCGTCTTGACAGA--                    763

Qy        767 AGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGC 826
                  | |||  |    ||||  | ||      | | |  |    |  |   ||
Db        764 ---GTACGTGGAGGTCGGCAGAAGTACAAGCGCAGGATAGATGCGGAGAACAGCCCA 817

Qy        827 CACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTC 886
              ||  |  ||    || ||  | ||       |    |       | ||    |   |
Db        818 TACCTGAACCCTCAGCTGGTTCAGCCAGCCAAAAAGCCATATAACAAGATTGTCTCACAT 877

Qy        887 ACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACC 946
                |  |||  |    |  |             ||       ||| | ||  |||  ||
Db        878 TTGTTGGTGGCTGAACCGGAGAAGATCTATGCCATGCCTGACCCTACTGTCCCCGACAGT 937

Qy        947 GAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATC 1006
              ||  |       | |  |||     | |||||||   |||||||    || ||
Db        938 GACATCAAAGCCCTCACTACACTGTGTGACTTGGCCGACCGAGAGTTGGTGGTTATCATT 997

Qy       1007 AGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTC 1066
               | ||||| ||| | ||||| |||||      |||   ||||   |||||  || | ||
Db        998 GGATGGGCGAAGCATATTCCAGGCTTCTCCACGCTGTCCCTGGCGGACCAGATGAGCCTT 1057

Qy       1067 TTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCAC 1126
               || ||||   |||||||||  | || ||  | ||  |  | |  || ||  ||
Db       1058 CTGCAGAGTGCTTGGATGGAAATTTTGATCCTTGGTGTCGTATACCGGTCTCTTTCATTT 1117

Qy       1127 CCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTA 1186
                |  | ||  ||| |||    ||| | |  |||||  ||  ||  ||||   ||
Db       1118 GAGGATGAACTTGTCTATGCAGACGATTATATAATGGACGAAGACCAGTCCAAAT--TA 1174

Qy       1187 GAAGGAATTCTGGAAATCTTTGACATGCTCCTGCCAACTACTTCAAGGTTTCGAGAGTTA 1246
                | |||  |||| ||  |   | |     |||||                | |      |
Db       1175 GCAGGCCTTCTTGATCTAAATAATGCTATCCTGCAGCTGGTAAAGAAATACAAGAGCATG 1234

Qy       1247 AAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTC 1299
              || ||  || |  |||||| | ||    |||| || ||    ||   |||||
Db       1235 AAGCTGGAAAAAGAAGAATTTGTCACCCTCAAAGCTATAGCTCTTGCTAATTC 1287


RESULT 18
US-09-041-886-10
; Sequence 10, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Shahrooz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  10:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3715 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  542..3286
US-09-041-886-10

  Query Match             5.0%;  Score 95.2;  DB 4;  Length 3715;
  Best Local Similarity   51.2%;  Pred. No. 2.8e-16;
  Matches  332;  Conservative    0;  Mismatches  298;  Indels   18;  Gaps    4;

Qy        520 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGG 579
              ||||    ||||  | ||| | || || || | ||||||||||      | |||| | |
Db       2202 CTGCCTGATCTGTCCACATGAAGCTTCTGGCTCTCACTATCGAGCTCTCACATGTCCAAG 2261

Qy        580 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGCACATAATGATTATATTTGTCCACCTAC 639
              || |||| ||| || |||||  | | |||  |  |  | ||  | ||  |    |
Db       2262 CTGCAAGGTCTTCTTCAAAACAGCCGCTGAAGGGAAACACAAGTACCTGTGCGCCAGCAG 2321

Qy        640 AAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 699
              |||| | || || || ||||||  |||  | || |  || |   | || || ||||||||
Db       2322 AAATGATTGCACTATTGATAAATTCCGAAGGAAAAATTGTCCATCTTGTCGTCTTCGGAA 2381

Qy        700 GTGTTACGAAGTGGGAATGGTCAAGTGTGGCTCCCGGAGAGAGACATGTGGGTACCGCCT 759
              ||||| ||||  || |||     | | | |     |    ||| | |||  | |
Db       2382 ATGTTATGAAGCAGGGATGACTCTGCGAGCCCCGAAGCTGAAGAAACTTGGTAATCTGAA 2441

Qy        760 TGTGCGCAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCCGCAAGCCAAGAGAAG 819
              | | | ||   || || | |  | |||| |  ||      | |    | |||    |
Db       2442 ACTACAC-GAGGAACCAGAGCCTTCCAGCACCACCAGCCCCACTCAGCACACAACCCACA 2500

Qy        820 TGGCGGCCAGCGCCCCCCAGTCGCGGGAGCTGCTGCTGCAGCGCTGAGCCCCAGCAGCT 879
              |  | |   |  | |     ||      ||          | |||||
Db       2501 AGCTGACAGTGTCACACATTGAAGCTATGAATG----------TCAGCCCATCTTTCTC 2550

Qy        880 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGCGCCCAGTGCGCC 939
              | ||  |      |    ||  |      ||||||  || | |  |  ||    ||
Db       2551 AATGTCCTGGAAGCCATTGAGCCAGGTGTACTGTGTGCTGGACA-CCACAACAACCAGCC 2609

Qy        940 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGCAGTTGATTACA 999
              |  | ||   ||  || || |      ||| |   |  |||  || |   || | |||||
Db       2610 CGACTCCTTTGCAGCCTTGCTCTCTAGCCTCAATGAACTCGGACAGACACAGCTTGTACA 2669

Qy       1000 CATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGT 1059
              | || |||  |||||||||    | || |||||      |  |   | ||  ||||||  |
Db       2670 CGTGGTCAAGTGGGCCAAGGCCTTGCCTGGCTTCCGCAACTTACACGTGGACGACCAGAT 2729

Qy       1060 GCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATCATGCGGCTCATGTGGCC------ 1113
              |    || |  ||  ||  |||||| | |  |  || | |    ||   |||||
Db       2730 GGCTGTCATTCAGTACTCCTGGATGGGGCTCATGCTCTTTGCCATGGCTGGCCATCCTT 2789

Qy       1114 CTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCT 1161
              | | | || | || || | | ||||  ||| || || |||||  || |
Db       2790 CACCAATGTCAACTCCAGGATGCTCTACTTCGCCCCTGATCTGGTTTT 2837


RESULT 19
US-08-076-726-15
; Sequence 15, Application US/08076726
; Patent No. 5464758
;  GENERAL INFORMATION:
;    APPLICANT:  Gossen, Manfred
;    APPLICANT:  Bujard, Hermann
;    TITLE OF INVENTION:  Tight Control of Gene Expression in
;    TITLE OF INVENTION:  Eucaryotic Cells by Tetracycline-responsive Promoters
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  District of Columbia
;      COUNTRY:  United States of America
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,726
;      FILING DATE:  14-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942.2490001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;      TELEX:  248636 SSK
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6244 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
US-08-076-726-15

  Query Match             4.7%;  Score 89;  DB 1;  Length 6244;
  Best Local Similarity   66.3%;  Pred. No. 1.8e-14;
  Matches  128;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        528 TCTCCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              ||||  | ||| | ||||| || | ||| ||||| |||    | ||||  | |||||||
Db       2365 TCTCCTGGCATCAAGCATCAGGCTGTCATTATGGTGTCCTCACCTGTGGCAGCTCTAAGG 2424

Qy        588 CCTTTTTAAAAGAAGCATTCAAGCACATAATGATTATATTTGTCCACCTACAAATCAGT 647
              ||| ||||||||    || |||| ||  || | ||| | ||| | |  | |||| | |
Db       2425 TCTTCTTTAAAAGGGCAATGCAAGGCCAGCATAACTATTTATGTGCTGGAAGAAATGACT 2484

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACG 707
              | |   | |||||| |||  | || | ||||| ||| || || ||| | ||||| |
Db       2485 GCATTGTTGATAAAATCCGCCAGGAAAAACTGCCCGGCGTGTCGCCTTAGAAAGTGCTGTC 2544

Qy        708 AAGTGGGAATCCT 720
              |||  || |||||
Db       2545 AAGCTGGCATGGT 2557


RESULT 20
US-08-260-452-8
; Sequence 8, Application US/08260452
; Patent No. 5650298
;  GENERAL INFORMATION:
;    APPLICANT:  Gossen, Manfred
;    APPLICANT:  Bujard, Hermann
;    APPLICANT:  Salfeld, Jochen
;    APPLICANT:  Voss, Jeffrey
;    TITLE OF INVENTION:  Tight Control of Gene Expression in Eucaryotic
;    TITLE OF INVENTION:  Cells by Tetracycline responsive Promoters
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive & Cockfield
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/260,452
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/076,427
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giulio A. DeConti, Jr.
;       REGISTRATION NUMBER:  31,503
;       REFERENCE/DOCKET NUMBER:  BBI 013CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227 7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6244 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;       STRAIN:  Towne (hCMV)
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR3
US-08-260-452-8

  Query Match              4.7%;  Score 89;  DB 1;  Length 6244;
  Best Local Similarity   66.3%;  Pred. No. 1.8e-14;
  Matches  128;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        528 TCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              |||| | ||| | ||||| || | ||| |||||| |||     | ||||    | |||||||
Db       2365 TCTGTGGGGATGAAGCATCAGGCTGTCATTATGGTGTCCTCACCTGTGGGAGCTGTAAGG 2424

Qy        588 CCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGT 647
               ||| ||||||||    ||  |||| ||  || | ||| | ||| | |  | |||| | |
Db       2425 TCTTCTTTAAAAGGGCAATGGAAGGGCAGCATAACTATTTATGTGCTCGAACAAATGACT 2484

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAACGTCTTACG 707
              | |   | |||||||| ||| | || | ||||| ||| || || ||| | ||||| |
Db       2485 GCATTGTTGATAAAATCCGCAGGAAAAACTGCCCGGCGTGTCGCCTTAGAAAGTGCTGTC 2544

Qy        708 AAGTGGGAATGGT 720
              ||| || |||||
Db       2545 AAGCTGGCATGGT 2557


RESULT 21
US-08-481-970-8
; Sequence 8, Application US/08481970
; Patent No. 5859310
;  GENERAL INFORMATION:
;     APPLICANT:  Gossen, Manfred
;     APPLICANT:  Bujard, Hermann
;     APPLICANT:  Salfeld, Jochen
;     APPLICANT:  Voss, Jeffrey
;     TITLE OF INVENTION:  Animal Transgenic for a Tetracycline Controlled    Transcriptl
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lahive & Cockfield
;       STREET:  60 State Street, Suite 510
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/481,970
;       FILING DATE:
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/260,452
;       FILING DATE:  14-JUN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/076,427
;       FILING DATE:  14-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giulio A. DeConti, Jr.
;       REGISTRATION NUMBER:  31,503
;       REFERENCE/DOCKET NUMBER:  BBI 013CP2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227 7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6244 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;       STRAIN:  Towne (hCMV)
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR3
US-08-481-970-8

  Query Match              4.7%;  Score 89;  DB 2;  Length 6244;
  Best Local Similarity   66.3%;  Pred. No. 1.8e-14;
  Matches  128;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        528 TCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              |||| | ||| | ||||| || | ||| |||||| |||     | ||||    | |||||||
Db       2365 TCTGTGGGGATGAAGCATCAGGCTGTCATTATGGTGTCCTCACCTGTGGGAGCTGTAAGG 2424

Qy        588 CCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGT 647
               ||| ||||||||    ||  |||| ||  || | ||| | ||| | |  | |||| | |
Db       2425 TCTTCTTTAAAAGGGCAATGGAAGGGCAGCATAACTATTTATGTGCTCGAACAAATGACT 2484

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAACGTCTTACG 707
              | |   | |||||||| ||| | || | ||||| ||| || || ||| | ||||| |
Db       2485 GCATTGTTGATAAAATCCGCAGGAAAAACTGCCCGGCGTGTCGCCTTAGAAAGTGCTGTC 2544

Qy        708 AAGTGGGAATGGT 720
              ||| || |||||
Db       2545 AAGCTGGCATGGT 2557


RESULT 22
US-08-897-719-8
; Sequence 8, Application US/08897719
; Patent No. 5922927
;  GENERAL INFORMATION:
;     APPLICANT:  Gossen, Manfred
;     APPLICANT:  Bujard, Hermann
;     APPLICANT:  Salfeld, Jochen
;     APPLICANT:  Voss, Jeffrey
;     TITLE OF INVENTION:  Tight Control of Gene Expression in Eucaryotic
;     TITLE OF INVENTION:  Cells by Tetracycline Responsive Promoters
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lahive & Cockfield
;       STREET:  60 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/897,719
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/260,452
;       FILING DATE:
;       APPLICATION NUMBER:  08/076,327
;       FILING DATE:  14-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giulio A. DeConti, Jr.
;       REGISTRATION NUMBER:  31,503
;       REFERENCE/DOCKET NUMBER:  BBI-013CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6244 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;       STRAIN:  Towne (hCMV)
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR3
US-08-897-719-8

  Query Match             4.7%;  Score 89;  DB 2;  Length 6244;
  Best Local Similarity   66.3%;  Pred. No. 1.8e-14;
  Matches  128;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        528 TCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              |||| | ||| | ||||| || | ||| ||||| ||| |  | ||||  | |||||||
Db       2365 TCTGTGGGGATGAAGCATCAGGCTGTCATTATGGTGTCCTCACCTGTGGCAGCTGTAAGG 2424

Qy        588 GCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGT 647
               ||| ||||||||  |  || |||| ||  || | ||| | ||| | |  | |||| | |
Db       2425 TCTTCTTTAAAAGGGCAATGGAAGGGCAGCATAACTATTTATGTGCTGGAAGAAATGACT 2484

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACG 707
              | |   | |||||||  |||  | || | |||||  ||| || || ||| | |||||  |
Db       2485 GCATTGTTGATAAAATCCGCAGGAAAAACTGCCCGGGCTGTCGCCTTAGAAAGTGCTGTC 2544

Qy        708 AAGTGGGAATGGT 720
              |||  || |||||
Db       2545 AAGCTGGCATGGT 2557


RESULT 23
US-09-163-269-8
; Sequence 8, Application US/09163269
; Patent No. 6252136
;  GENERAL INFORMATION:
;     APPLICANT:  Gossen, Manfred
;     APPLICANT:  Bujard, Hermann
;     APPLICANT:  Salfeld, Jochen
;     APPLICANT:  Voss, Jeffrey
;     TITLE OF INVENTION:  Animal Transgenic for a Tetracycline-Controlled    Transcripti
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lahive & Cockfield
;       STREET:  60 State Street, Suite 510
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/163,269
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/481,970
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/076,327
;       FILING DATE:  14-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giulio A. DeConti, Jr.
;       REGISTRATION NUMBER:  31,503
;       REFERENCE/DOCKET NUMBER:  BBI-013CP2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6244 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Human cytomegalovirus
;       STRAIN:  Towne (hCMV)
;     IMMEDIATE SOURCE:
;       CLONE:  pUHD BGR3
US-09-163-269-8

  Query Match             4.7%;  Score 89;  DB 4;  Length 6244;
  Best Local Similarity   66.3%;  Pred. No. 1.8e-14;
  Matches  128;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        528 TCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              |||| | ||| | ||||| || | ||| ||||| ||| |  | ||||  | |||||||
Db       2365 TCTGTGGGGATGAAGCATCAGGCTGTCATTATGGTGTCCTCACCTGTGGCAGCTGTAAGG 2424

Qy        588 GCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGT 647
               ||| ||||||||  |  || |||| ||  || | ||| | ||| | |  | |||| | |
Db       2425 TCTTCTTTAAAAGGGCAATGGAAGGGCAGCATAACTATTTATGTGCTGGAAGAAATGACT 2484

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACG 707
              | |   | |||||||  |||  | || | |||||  ||| || || ||| | |||||  |
Db       2485 GCATTGTTGATAAAATCCGCAGGAAAAACTGCCCGGGCTGTCGCCTTAGAAAGTGCTGTC 2544

Qy        708 AAGTGGGAATGGT 720
              |||  || |||||
Db       2545 AAGCTGGCATGGT 2557


RESULT 24
US-08-479-914E-1
; Sequence 1, Application US/08479914E
; Patent No. 6416998
; GENERAL INFORMATION:
;  APPLICANT: O'Malley, Bert W.
;  APPLICANT:  Tsai, Ming-Jer
;  APPLICANT:  Ledebur, Harry C.
;  APPLICANT:  Kittle, Joseph D.
;  TITLE OF INVENTION: MODIFIED STEROID HORMONES FOR GENE THERAPY AND METHODS FOR THE
;  FILE REFERENCE: 212/133
;  CURRENT APPLICATION NUMBER: US/08/479,914E
;  CURRENT FILING DATE:  1995-06-07
;  PRIOR APPLICATION NUMBER: US 07/939,246
;  PRIOR FILING DATE: 1992-09-02
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: MicroSoft Word 97
; SEQ ID NO 1
```

```
;   LENGTH: 6177
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Plasmid pGR0403R
US-08-479-913E-1

  Query Match              4.5%;  Score 84.8;  DB 4;  Length 6177;
  Best Local Similarity   64.0%;  Pred. No. 2.5e-14;
  Matches  128;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        519 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 578
              ||||| || ||| |||| | || || |||| ||| || || ||   | | |||| |
Db       5815 TCTGCCTGGTGTGCTCCGATGAAGCTTCAGGATGTCATTACGGGGTGCTGACATGTGGAA 5874

Qy        579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638
              | || || |  || ||||||||    | |||||||  |  ||||  ||||| | |  |
Db       5875 GCTGCAAAGTATTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 5934

Qy        639 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 698
              |||  | || |  || |||||||  ||  | || | ||||  || |||||   |||||
Db       5935 GAAACGATTGCATCATTGATAAAATTCGAAGGAAAAACTGCCCAGCATGCCGCTATCGGA 5994

Qy        699 AGTGTTACGAAGTGGGAATG 718
              | |||    | |  ||||||
Db       5995 AATGTCTTCAGGCTGGAATG 6014


RESULT 25
US-09-091-042A-1
; Sequence 1, Application US/09091042A
; Patent No. 6455600
;   GENERAL INFORMATION:
;        APPLICANT: The Government of the United States of America
;                   as respresented by the Secretary
;                   Department of Health and Human Services
;                   Washington, D.C.
;                   Hunn Ph.D., Han
;                   Hager Ph.D., Gordon L.
;        TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MONITORING
;                            DNA BINDING MOLECULES IN LIVING CELLS
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Needle & Rosenberg
;             STREET: 127 Peachtree Street, Suite 1200
;             CITY: Atlanta
;             STATE: Georgia
;             COUNTRY: USA
;             ZIP: 30303
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/091,042A
;             FILING DATE: 08-Jun-1998
;             CLASSIFICATION: -Unknown-
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/008,473
;             FILING DATE: 08 Dec 1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Selby, Elizabeth
;             REGISTRATION NUMBER: 38298
;             REFERENCE/DOCKET NUMBER: 14014.0183
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 404-688 0770
;             TELEFAX: 404-688 9880
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 7257 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: nucleic acid
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1072..4284
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-091-042A-1

  Query Match              4.5%;  Score 84.8;  DB 4;  Length 7257;
  Best Local Similarity   64.0%;  Pred. No. 2.7e-14;
  Matches  128;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        519 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 578
              ||||| || ||| |||| | || || |||| ||| || || ||   | | |||| |
Db       3212 TCTGCCTGGTGTGCTCCGATGAAGCTTCAGGATGTCATTACGGGGTGCTGACATGTGGAA 3271

Qy        579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638
              | || || |  || ||||||||    | |||||||  |  ||||  ||||| | |  |
Db       3272 GCTGCAAAGTATTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 3331

Qy        639 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 698
              |||  | || |  || |||||||  ||  | || | ||||  || |||||   |||||
Db       3332 GAAACGATTGCATCATTGATAAAATTCGAAGGAAAAACTGCCCAGCATGCCGCTATCGGA 3391

Qy        699 AGTGTTACGAAGTGGGAATG 718
              | |||    | |  ||||||
Db       3392 AATGTCTTCAGGCTGGAATG 3411


RESULT 26
US-08-629-939-1
; Sequence 1, Application US/08629939
; Patent No. 5645995
;  GENERAL INFORMATION:
;    APPLICANT:  Kieback, Dirk G.
;    TITLE OF INVENTION:  METHODS FOR DIAGNOSING AN INCREASED
;    TITLE OF INVENTION:  RISK OF BREAST OR OVARIAN CANCER
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W., Suite 800
;      CITY:  Washington, D.C.
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,939
;      FILING DATE:  12-APRIL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KIT, Gordon
;      REGISTRATION NUMBER:  30,764
;      REFERENCE/DOCKET NUMBER:  A-6612
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293 7060
;      TELEFAX:  (202) 293 7860
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3014 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
US-08-629-939-1

  Query Match              4.3%;  Score 81;  DB 1;  Length 3014;
```

```
  Best Local Similarity   63.7%;  Pred. No. 2e-12;
  Matches  123;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    1;

Qy        528 TCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              ||||  | ||| | ||||| || | ||| ||||| |||     | ||||   | |||||||
Db       1881 TCTGTGGGGATGAAGCATCAGGCTGTCATTATGGTGTCCTTACCTGTGGCAGCTGTAAGG 1940

Qy        588 CCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGT 647
              ||| ||||| ||    ||  |||| ||  |  | ||  | ||| | |  | |||| | |
Db       1941 TCTTCTTTAAGAGGGCAATGGAAGGGCAGCACAACTACTTATGTGCTGGAAGAAATGACT 2000

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACG 707
              | |   | ||||||| |||  | || | |||||  || || || ||| | |||||
Db       2001 GCATCGTTGATAAAATCCGCAGAAAAAACTGCCCAGCATGTCGCCTTAGAAAGTGCTGTC 2060

Qy        708 AAGTGGGAATGGT 720
              | |  || |||||
Db       2061 AGGCTGGCATGGT 2073


RESULT 27
US-08-759-873-1
; Sequence 1, Application US/08759873
; Patent No. 5683885
;  GENERAL INFORMATION:
;    APPLICANT:  Kioback, Dirk G.
;    TITLE OF INVENTION:  METHODS FOR DIAGNOSING AN INCREASED RISK
;    TITLE OF INVENTION:  OF BREAST OR OVARIAN CANCER
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W., Suite 800
;      CITY:  Washington, D.C.
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/759,873
;      FILING DATE:  12-APRIL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KII, Gordon
;      REGISTRATION NUMBER:  30,764
;      REFERENCE/DOCKET NUMBER:  A-6612
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3014 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
US-08-759-873-1

  Query Match             4.3%;  Score 81;  DB 1;  Length 3014;
  Best Local Similarity   63.7%;  Pred. No. 2e-12;
  Matches  123;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        528 TCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              ||||  | ||| | ||||| || | ||| ||||| |||     | ||||   | |||||||
Db       1881 TCTGTGGGGATGAAGCATCAGGCTGTCATTATGGTGTCCTTACCTGTGGCAGCTGTAAGG 1940

Qy        588 CCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGT 647
              ||| ||||| ||    ||  |||| ||  |  | ||  | ||| | |  | |||| | |
Db       1941 TCTTCTTTAAGAGGGCAATGGAAGGGCAGCACAACTACTTATGTGCTGGAAGAAATGACT 2000

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACG 707
              | |   | ||||||| |||  | || | |||||  || || || ||| | |||||
Db       2001 GCATCGTTGATAAAATCCGCAGAAAAAACTGCCCAGCATGTCGCCTTAGAAAGTGCTGTC 2060

Qy        708 AAGTGGGAATGGT 720
              | |  || |||||
Db       2061 AGGCTGGCATGGT 2073


RESULT 28
5260432-1
;Patent No. 5260432
;    APPLICANT: TAKAKU, FUMIMARO;ISHIKAWA, TAKASHI;IMAWARI, MICHIO;
;EVANS, RONALD M.;UMESONO, KAZUHIKO
;    TITLE OF INVENTION: HUMAN GAMMA RETINOIC ACID RECEPTOR DNA
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/370,407
;      FILING DATE: 22-JUN-1989
;SEQ ID NO:1:
;      LENGTH: 1576
5260432-1

  Query Match             4.3%;  Score 80.8;  DB 6;  Length 1576;
  Best Local Similarity   62.3%;  Pred. No. 1.7e-12;
  Matches  127;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        521 TGCGCTGGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
              |||   || ||||  ||  |  | || || || |||||||| ||| | || ||||||||
Db        467 TGCTTCGTGTGCAATGACAAGTCCTCTGGCTACCACTATGCGGTCAGCTCTTGTCAAGGC 526

Qy        581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
              || |||| ||| |||    ||||||| ||     | |  |  || |  ||||
Db        527 TGCAAGGGCTTCTTTCGCCGAAGCATCCAGAAGAACATGGTGTACACGTGTCACCGCGAC 586

Qy        641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
              ||  | ||||  ||| | ||        | ||  ||||||||  |||||| || | ||||
Db        587 AAAAACTGTATCATCAACAAGGTGACCAGGAATCGCTGCCAGTACTGCCGGCTACAGAAG 646

Qy        701 TGTTACGAAGTGGGCAATCGTCAAG 724
              || | |||||||| |||   |||
Db        647 TGCTTCGAAGTGGGCATGTCCAAG 670


RESULT 29
5223606-1
;Patent No. 5223606
;    APPLICANT:  BLAUDIN DE THE, HUGUES;MARCHIO, AGNES;TIOLLAIS,
;PIERRE;DEJEAN, ANNE
;    TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED
;PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA
;    NUMBER OF SEQUENCES: 11
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/134,130
;      FILING DATE: 17-DEC-1987
;    PRIOR APPLICATION DATA:
;SEQ ID NO:1:
;      LENGTH: 2989
5223606-1

  Query Match             4.1%;  Score 78;  DB 6;  Length 2989;
  Best Local Similarity   61.9%;  Pred. No. 1.3e-11;
  Matches  140;  Conservative    0;  Mismatches   85;  Indels    1;  Gaps    1;

Qy        520 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGG 579
              ||||   ||||||   ||  |  |||| || || |||||||| ||| |  | ||||| ||
Db        560 CTGCTTCGTCTGCCAGGACAAATCATCAGGGTACCACTATGGGGTCAGCGCCTGTGAGGG 619

Qy        580 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 639
              |||||||| ||||||   ||||| |||||    |||   ||| | ||||| ||
Db        620 ATGTAAGGGCTTTTTCCGCAGAAGTATTCAGAAGAATATGATTTACACTTGTCACCGAGA 679
```

```
Qy          640 AAATCAGTGTACAATCGATAAAAACCGGC-GCAAGAGCTGCCAGGCCTGCCGACTTCGGA 698
                ||  | |||   ||  |||||  |  | | ||  | |||||   ||| ||||| | ||
Db          680 TAAGAACTGTGTTATTAATAAAAGTCACCAGGAATCGATGCCAATACTGTCCGACTCCAGA 739

Qy          699 AGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAG 744
                |||| |  |||||||||.||   ||         | |||  || ||
Db          740 AGTGCTTTGAAGTGGGAATGTCCAAAGAATCTGTCAGGAATGACAG 785


RESULT 30
US-09-277-078-1
; Sequence 1, Application US/09277078
; Patent No. 6312949
; GENERAL INFORMATION:
;  APPLICANT: Sakurada, Kazuhiro
;  APPLICANT: Palmer, Theo
;  APPLICANT: Gage, Fred H.
;  TITLE OF INVENTION: REGULATION OF TYROSINE HYDROXYLASE
;  TITLE OF INVENTION: EXPRESSION
;  FILE REFERENCE: 07251/031001
;  CURRENT APPLICATION NUMBER: US/09/277,078
;  CURRENT FILING DATE: 1999-03-26
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1797
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(900)
US-09-277-078-1

  Query Match             3.9%;  Score 74.4;  DB 4;  Length 1797;
  Best Local Similarity   57.9%;  Pred. No. 1e-15;
  Matches  142;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy          519 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 578
                | |||||||| ||| | ||  ||||  | |  | || || || ||  |  | |||||
Db          785 TGTGCGCTGTTTGCGGTGACAACGCGGCCTGTCAGCATTACGGTGTTCGCACTTGTGAGG 844

Qy          579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638
                | || || |  || |||||  | |   | |||  | |      | ||| | |||  ||| |
Db          845 GCTGCAAAGGTTTCTTTAAGCGCACGGTGCAAAAAAACGCGAAATATGTGTGTTTAGCAA 904

Qy          639 CAAATCAGTGTACAATCGATAAAAACCGGCGCAACAGCTGCCAGGCCTGCCGACTTCGGA 698
                  ||  | ||  || | |||||   ||| || ||  | || |||  ||| ||  ||| ||
Db          905 ATAAAAATGCCCAGTGGATAAGCGCCGCCGAAATCGTTGTCAGTACTGTCGGTTTCAGA 964

Qy          699 AGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGACAT 746
                ||||    |  || || |||| ||    |     || |  || || |
Db          965 AGTGCCTGGCTGTTGGGATGGTTAAAGAAGTGGTTCGCACGGACAGTT 1012


RESULT 41
US-08-592-484-3
; Sequence 3, Application US/08592484
; Patent No. 5830760
;  GENERAL INFORMATION:
;     APPLICANT:  Tsai, S. and S.J. Collins
;     TITLE OF INVENTION:  "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Recept
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;       STREET:  2800 Pacific First Centre, 1420 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98101
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette 5.25 inch, 1.2Mb storage
;       COMPUTER:  IBM PC/486 Compatible
;       OPERATING SYSTEM:  MS DOS 4.01
;       SOFTWARE:  Word for Windows 5.01-1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/592,484
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/099,242
;       FILING DATE:  July 28, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Broderick, Thomas F.
;       REGISTRATION NUMBER:  31,832
;       REFERENCE/DOCKET NUMBER:  FHCR-1-7190
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206 682 8190; 1-206-224 0709(direct)
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2658 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;       DESCRIPTION:  page 4,RAR-alpha403 dominant negative; deleted of 1311-1596 of S
;     IMMEDIATE SOURCE:
;       LIBRARY:  cDNA
US-08-592-484-3

  Query Match             3.9%;  Score 74.4;  DB 2;  Length 2658;
  Best Local Similarity   60.3%;  Pred. No. 1.2e-10;
  Matches  124;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy          521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
                |||  ||||||    ||  |  | || || || |||||||| ||| |  | |||||  ||
Db          364 TGCTTTGTCTGTCAGGACAAGTCCTCAGCCTACCACTATGCGGTCAGCGCCTGTGAGGGC 423

Qy          581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
                || |||| ||| ||     | |||||  ||     | |  |  || |  ||||
Db          424 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 483

Qy          641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
                ||  | ||  |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db          484 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 543

Qy          701 TGTTACGAAGTGGGAATGGTGAAG 724
                || |  |||||||| |||   |||
Db          544 TGCTTTGAAGTGGGCATGCCCAAG 567


RESULT 42
US-08-095-728B-3
; Sequence 3, Application US/08095728B
; Patent No. 5843642
;  GENERAL INFORMATION:
;     APPLICANT:  DMITROVSKY, ETHAN
;     APPLICANT:  WARRELL JR, RAYMOND P
;     APPLICANT:  MILLER JR, WILSON H
;     APPLICANT:  FRANKEL, STANLEY
;     TITLE OF INVENTION:  METHODS FOR THE DETECTION AND
;     TITLE OF INVENTION:  TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)
;     NUMBER OF SEQUENCES:  11
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  COOPER & DUNHAM LLP
;       STREET:  1185 AVENUE OF THE AMERICAS
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  U.S.A.
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.24
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/095,728B
```

```
;       FILING DATE:  21-JUL-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/673,838
;       FILING DATE:  22-MAR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  WHITE, JOHN P
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  38694-A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278-0400
;       TELEFAX:  (212) 391-0525
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2928 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  hRAR ALPHA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  103..1488
;       OTHER INFORMATION:
US-08-095-728B-3

  Query Match              3.9%;  Score 74.4;  DB 2;  Length 2928;
  Best Local Similarity   60.3%;  Pred. No. 1.3e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTCGTCGAAGGA 580
              ||| |||||| |  || | | || || || |||||||| ||| | | ||||| ||
Db        364 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 423

Qy        581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
              || |||| ||| ||     | ||||| ||     | | |  || | ||||
Db        424 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 483

Qy        641 AATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
              ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db        484 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 543

Qy        701 TGTTACGAAGTGGGAATGGTGAAG 724
              || |  |||||||| |||   |||
Db        544 TGCTTTGAAGTGGGCATGTCCAAG 567


RESULT 33
PCT-US92-02320A-3
; Sequence 3, Application PC/TUS92023?2A
;  GENERAL INFORMATION:
;    APPLICANT:  Sloan-Kettering Institute, For Cancer Research
;    TITLE OF INVENTION:  METHODS FOR DETECTION AND TREATMENT OF CANCER
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOPER & DUNHAM
;      STREET:  30 ROCKEFELLER PLAZA
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/02320A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 675,084
;      FILING DATE:  22-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 673,848
;      FILING DATE:  22-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WHITE, JOHN P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  38694-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 644-0525
;      TELEX:  (212) 422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  hRAR ALPHA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1488
;      OTHER INFORMATION:
PCT US92 02320A-3

  Query Match              3.9%;  Score 74.4;  DB 5;  Length 2928;
  Best Local Similarity   60.3%;  Pred. No. 1.3e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTCGTCGAAGGA 580
              ||| |||||| |  || | | || || || |||||||| ||| | | ||||| ||
Db        364 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 423

Qy        581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
              || |||| ||| ||     | ||||| ||     | | |  || | ||||
Db        424 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 483

Qy        641 AATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
              ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db        484 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 543

Qy        701 TGTTACGAAGTGGGAATGCTGAAC 724
              || |  |||||||| |||   |||
Db        544 TGCTTTGAAGTGGGCATGTCCAAG 567


RESULT 34
US-08-592-383-1
; Sequence 1, Application US/08592383
; Patent No. 5830760
;  GENERAL INFORMATION:
;    APPLICANT:  Tsai, S. and S.J. Collins
;    TITLE OF INVENTION:  "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Rec
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Centre, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-5.25 inch, 1.2Mb storage
;      COMPUTER:  IBM PC/386 Compatible
;      OPERATING SYSTEM:  MS-DOS 4.01
;      SOFTWARE:  Word for Windows 5.01-t
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,383
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  08/099,242
;       FILING DATE:  July 28, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Broderick, Thomas F.
;       REGISTRATION NUMBER:  31,832
;       REFERENCE/DOCKET NUMBER:  FHCR-1-7190
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1 206 682 8100; 1-206-224 0709(direct)
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2940 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;       DESCRIPTION:  see Figure 16B; RAR alpha
;     IMMEDIATE SOURCE:
;       LIBRARY:  cDNA
US-08-592-384-1

  Query Match                 3.9%;  Score 74.4;  DB 2;  Length 2940;
  Best Local Similarity   60.3%;  Pred. No. 1.3e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
               ||| |||||| |  || | | | || || || |||||||| ||| | | | ||||| ||
Db         364 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 423

Qy         581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
               || |||| ||| ||     | ||||| ||     | |  |  || |  ||||
Db         424 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 483

Qy         641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
               ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db         484 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 543

Qy         701 TGTTACGAAGTGGGAATGGTGAAG 724
               || |  |||||||| |||   |||
Db         544 TGCTTTGAAGTGGGCATGTCCAAG 567


RESULT 45
5171671-1
;Patent No. 5171671
;     APPLICANT: EVANS, RONALD M.;ONG, ESTELITA S.;SEGUI,
;PRUDIMAR S.;THOMPSON, CATHERINE C.;UEMSONO, KAZUHIKO
;GUGUERE, VINCENT
;     TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION
;     NUMBER OF SEQUENCES: 2
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/546,256
;       FILING DATE: 06 AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 276,536
;       FILING DATE: 30 NOV-1988
;       APPLICATION NUMBER: 128,441
;       FILING DATE: 02 DEC-1987
;SEQ ID NO:1:
;       LENGTH: 2940
5171671-1

  Query Match                 3.9%;  Score 74.4;  DB 6;  Length 2940;
  Best Local Similarity   60.3%;  Pred. No. 1.3e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
               ||| |||||| |  || | | | || || || |||||||| ||| | | | ||||| ||
Db         364 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 423

Qy         581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
               || |||| ||| ||     | ||||| ||     | |  |  || |  ||||
Db         424 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 483

Qy         641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
               ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db         484 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 543

Qy         701 TGTTACGAAGTGGGAATGGTGAAG 724
               || |  |||||||| |||   |||
Db         544 TGCTTTGAAGTGGGCATGTCCAAG 567


RESULT 46
US-08-306-691B-52
; Sequence 52, Application US/08306691B
; Patent No. 5734039
;  GENERAL INFORMATION:
;     APPLICANT:  Calabretta, Bruno
;     APPLICANT:  Skorski, Tomasz
;     TITLE OF INVENTION:  ANTISENSE
;     TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;       STREET:  Two Penn Center, Suite 1800
;       CITY:  Philadelphia
;       STATE:  Pennsylvania
;       COUNTRY:  U.S.A.
;       ZIP:  19102
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/306,691B
;       FILING DATE:  September 15, 1994
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Monaco, Daniel A.
;       REGISTRATION NUMBER:  30,480
;       REFERENCE/DOCKET NUMBER:  8321-8
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (215) 568-8383
;       TELEFAX:  (215) 568-5549
;       TELEX:  No. 5734039e
;   INFORMATION FOR SEQ ID NO:  52:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3036 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-306-691B-52

  Query Match                 3.9%;  Score 74.4;  DB 1;  Length 3036;
  Best Local Similarity   60.3%;  Pred. No. 1.4e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
               ||| |||||| |  || | | | || || || |||||||| ||| | | | ||||| ||
Db        1333 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 1392

Qy         581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
               || |||| ||| ||     | ||||| ||     | |  |  || |  ||||
Db        1393 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 1452

Qy         641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
               ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db        1453 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 1512

Qy         701 TGTTACGAAGTGGGAATGGTGAAG 724
               || |  |||||||| |||   |||
Db        1513 TGCTTTGAAGTGGGCATGTCCAAG 1536
```

```
RESULT 37
US-08-095-728B-1
; Sequence 1, Application US/08095728B
; Patent No. 5843642
;  GENERAL INFORMATION:
;    APPLICANT:  DMITROVSKY, ETHAN
;    APPLICANT:  WARRELL JR, RAYMOND P
;    APPLICANT:  MILLER JR, WILSON H
;    APPLICANT:  FRANKEL, STANLEY
;    TITLE OF INVENTION:  METHODS FOR THE DETECTION AND
;    TITLE OF INVENTION:  TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOPER & DUNHAM LLP
;      STREET:  1185 AVENUE OF THE AMERICAS
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/095,728B
;      FILING DATE:  21-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/673,838
;      FILING DATE:  22-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WHITE, JOHN P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  38694-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3036 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  MYL-RAR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  67..2457
;      OTHER INFORMATION:
US-08-095-728B-1

  Query Match             3.9%;  Score 74.4;  DB 2;  Length 3036;
  Best Local Similarity   60.3%;  Pred. No. 1.3e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
              ||| |||||| || | | || || || ||||||| ||| | | ||||| ||
Db       1333 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 1392

Qy        581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
              || |||| ||| ||     | ||||| ||     | |  |  || | |  ||||
Db       1393 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 1452

Qy        641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
              ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db       1453 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCACAAG 1512

Qy        701 TGTTACGAAGTGGGAATGGTGAAG 724
              || |  |||||||| |||  |||
Db       1513 TGCTTTGAAGTGGGCATGTCCAAG 1536


RESULT 38
PCT-US92-02320A-1
; Sequence 1, Application PC/TUS9202320A
;  GENERAL INFORMATION:
;    APPLICANT:  Sloan-Kettering Institute, For Cancer Research
;    TITLE OF INVENTION:  METHODS FOR DETECTION AND TREATMENT OF CANCER
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOPER & DUNHAM
;      STREET:  30 ROCKEFELLER PLAZA
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/02320A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 675,084
;      FILING DATE:  22-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 673,838
;      FILING DATE:  22-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WHITE, JOHN P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  38694-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 644-0525
;      TELEX:  (212) 422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3036 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  MYL-RAR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  67..2457
;      OTHER INFORMATION:
PCT-US92-02320A-1

  Query Match             3.9%;  Score 74.4;  DB 5;  Length 3036;
  Best Local Similarity   60.3%;  Pred. No. 1.3e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
              ||| |||||| || | | || || || ||||||| ||| | | ||||| ||
Db       1333 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 1392

Qy        581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
              || |||| ||| ||     | ||||| ||     | |  |  || | |  ||||
Db       1393 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 1452

Qy        641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
              ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db       1453 AACAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 1512

Qy        701 TGTTACGAAGTGGGAATGGTGAAG 724
              || |  |||||||| |||  |||
Db       1513 TGCTTTGAAGTGGGCATGTCCAAG 1536
```

```
Db        1513 TGCTTTGAAGTGGGCATGTCCAAG 1536


RESULT 39
US-08-312-312A-1/c
; Sequence 1, Application US/08312312A
; Patent No. 5550024
;  GENERAL INFORMATION:
;    APPLICANT:  Rothschild, Max F.
;    APPLICANT:  Fuqqle, Christopher K.
;    APPLICANT:  Jacobson, Carol D.
;    APPLICANT:  Mileham, Alan J.
;    APPLICANT:  Plastow, Graham S.
;    APPLICANT:  Vaske, David A.
;    TITLE OF INVENTION:  Genetic Markers for Pig Litter Size
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zarley, McKee, Thomte, Voorhees & Sease
;      STREET:  801 Grand Suite 3200
;      CITY:  Des Moines
;      STATE:  Iowa
;      COUNTRY:  United States
;      ZIP:  50309
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/312,312A
;      FILING DATE:  26-SEP 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/961,819
;      FILING DATE:  19-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/687,708
;      FILING DATE:  19-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nebel, Heidi S.
;      REGISTRATION NUMBER:  37,719
;      REFERENCE/DOCKET NUMBER:  BRDC-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  515-288-3667
;      TELEFAX:  515 288 1338
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4460 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  303..304
;      OTHER INFORMATION:  /note= "Gap in DNA sequence of
;      OTHER INFORMATION:  undetermined length"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1197..1198
;      OTHER INFORMATION:  /note= "Gap in DNA sequence of
;      OTHER INFORMATION:  undetermined length"
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  534..600
US-08-312-312A-1

  Query Match              3.9%;  Score 74.4;  DB 1;  Length 3460;
  Best Local Similarity   81.7%;  Pred. No. 1.4e-10;
  Matches   98;  Conservative    0;  Mismatches   21;  Indels    1;  Gaps    1;

Qy         610 AGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCG 669
               ||||||||||||| || || |||||||||||| || ||||| ||||| ||||| ||| || |
Db         602 AGGACATAATGACTACATGTGTCCAGCTACCAACCAGTCCACAATTGATAAGAACAGGAG 543

Qy         670 CAAGACGCTGCCAGCGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 729
               |||||||| || |||||||| || || ||||| |||||||||||| ||| |||| | ||
Db         542 GAAGAGCTGTCA-GCCTGCCGGCTACGCAAGTGCTACGAAGTGGGCATGATGAAAGGGGG 484


RESULT 40
US-08-892-747-13
; Sequence 13, Application US/08892747
; Patent No. 6057153
;  GENERAL INFORMATION:
;    APPLICANT:  Shaji T. George, Michael Ma, Martina Werner,
;    APPLICANT:  Umberto Pace and Allan R. Goldberg
;    TITLE OF INVENTION:  Stabilized External Guide Sequences
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/892,747
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/372,556
;      FILING DATE:  January 13, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00513
;      FILING DATE:  January 19, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  ILI109C1P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..4511
;      OTHER INFORMATION:  /function= "PML-RAR' DNA Sequence."
US-08-892-747-13

  Query Match              3.9%;  Score 74.4;  DB 3;  Length 4511;
  Best Local Similarity   60.8%;  Pred. No. 1.4e-10;
  Matches  123;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         521 TGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGA 580
               ||| |||||| |  || | | || || || |||||||| ||| | | ||||| ||
Db        1808 TGCTTTGTCTGTCAGGACAAGTCCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGGC 1867
```

```
Qy        581 TGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACA 640
              || |||| ||| ||     | ||||| ||     | |  |  || |  ||||
Db       1868 TGCAAGGGCTTCTTCCGCCGCAGCATCCAGAAGAACATGGTGTACACGTGTCACCGGGAC 1927

Qy        641 AATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAG 700
              ||  | || |  ||| | ||       || ||  |||||||  |||||||| | ||||
Db       1928 AAGAACTGCATCATCAACAAGGTGACCCGGAACCGCTGCCAGTACTGCCGACTGCAGAAG 1987

Qy        701 TGTTACGAAGTGGGAATGGTGAAG 724
              || |  |||||||| |||   |||
Db       1988 TGCTTTCAAGTGGGCATGTCCAAG 2011


Search completed: June 16, 2003, 03:20:55
Job time : 94.375 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 03:16:17 ; Search time 283.725 Seconds
(without alignments)
9686.913 Million cell updates/sec

Title: US-08-826-361A-24
Perfect score: 1898
Sequence: 1 CACGAATCTTTGAGAACATT ... TGAATGGAGGGTTTGCACAC 1898

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCTUS_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1718	90.5	1740	9	US-10-198-785-1	Sequence 1, Appli
2	1718	90.5	1740	10	US-09-815-068A-1	Sequence 1, Appli
3	438	23.1	597	10	US-09-864-761-7903	Sequence 7903, Ap
4	395.8	20.9	2764	10	US-09-893-666A-1	Sequence 1, Appli
5	390.8	20.6	1788	10	US-09-853-033-1	Sequence 1, Appli
6	390.8	20.6	4963	10	US-09-281-674-9	Sequence 9, Appli
7	390.8	20.6	4963	10	US-09-892-227-9	Sequence 9, Appli
8	389.2	20.5	2092	9	US-10-052-092-6	Sequence 6, Appli
9	389.2	20.5	2092	9	US 09 952-680A-9	Sequence 9, Appli
10	389.2	20.5	6450	9	US-10-096 710 2	Sequence 2, Appli
11	389.2	20.5	6450	9	US-10-081-563-1	Sequence 1, Appli
12	389.2	20.5	6450	9	US 10 052 092 1	Sequence 1, Appli
13	389.2	20.5	6450	9	US-10-052 092 7	Sequence 7, Appli
14	389.2	20.5	6610	9	US-10-052-092-20	Sequence 20, Appl
15	388	20.4	1788	9	US-10-052-092-24	Sequence 24, Appl
16	385.2	20.3	6698	9	US 10 198-846 11205	Sequence 11205, A
17	366.2	19.3	1978	9	US-10-052 092 23	Sequence 23, Appl
18	326.6	17.2	2106	9	US-10-052-092-5	Sequence 5, Appli
19	306.8	16.2	3300	9	US-10-259-864-3	Sequence 3, Appli
20	306	16.1	306	10	US-09-864-761-24607	Sequence 24607, A
21	302	15.9	586	10	US-09-864-761-16388	Sequence 16388, A
22	300	15.8	300	10	US 09 864 761 32872	Sequence 32872, A
23	285.6	15.0	1644	9	US-10 052 092 2	Sequence 2, Appli
24	285.6	15.0	11522	9	US-10-052-092-19	Sequence 19, Appl
25	284	15.0	1983	10	US-09-853-033-3	Sequence 3, Appli
26	280.8	14.8	1983	10	US-09-853-033-7	Sequence 7, Appli
27	279.2	14.7	1983	10	US-09-853-033-5	Sequence 5, Appli
28	278	14.6	451	9	US-09-918-995-15134	Sequence 15134, A
29	267	14.1	973	10	US-09-737-255-4	Sequence 4, Appli
30	265.4	14.0	2739	9	US-09-908-153B-39	Sequence 39, Appl
31	265.4	14.0	2739	9	US-09-908 153B-44	Sequence 44, Appl
32	265.4	14.0	2775	9	US-09-908-153B-41	Sequence 41, Appl
33	265.4	14.0	2775	9	US-09-908-153B-43	Sequence 43, Appl
34	232.4	12.2	599	10	US-09 864 761 6718	Sequence 6718, Ap
35	183	9.6	183	10	US-09-864-761-23452	Sequence 23452, A
36	133.8	7.0	2218	10	US 09 962 436 268	Sequence 268, App
37	129	6.8	2402	9	US 10 211 239 7	Sequence 7, Appli
38	129	6.8	2539	9	US-10-037-270-454	Sequence 454, App
39	97.2	5.1	3577	9	US-10-008-739A-1	Sequence 1, Appli
40	95.2	5.0	1329	9	US-09-997-267-1	Sequence 1, Appli
41	95.2	5.0	3715	10	US-09-880-107-2300	Sequence 2300, Ap
42	95.2	5.0	4393	9	US-10-198-846-10819	Sequence 10819, A
43	95	5.0	2582	9	US 10 202-846 1	Sequence 1, Appli
44	93.6	4.9	1171	9	US-09-997-267-3	Sequence 3, Appli
45	89	4.7	6244	10	US-09-281-674-8	Sequence 8, Appli
46	89	4.7	6244	10	US-09-892-227-8	Sequence 8, Appli
47	88.6	4.7	2481	9	US-10-211-239-8	Sequence 8, Appli
48	88.6	4.7	2481	10	US-09-853-386-89	Sequence 89, Appl
49	88.6	4.7	2626	12	US 10 002 600 45	Sequence 45, Appl
50	88.4	4.7	2629	10	US 09 964 456 1892	Sequence 1892, Ap
51	88.4	4.7	2750	12	US-10-002-600 44	Sequence 44, Appl
52	88	4.6	465247	10	US 09 944 267A 1	Sequence 1, Appli
53	86.8	4.6	600	9	US-10 153 668 219	Sequence 219, App
54	86.4	4.6	920	9	US 10 153 668 221	Sequence 221, App
55	86.4	4.6	1881	9	US-10-153-668-223	Sequence 223, App
56	86.4	4.6	4789	9	US 10 153 668 225	Sequence 225, App
57	84.8	4.5	1264	9	US-10-153-668-217	Sequence 217, App
58	84.8	4.5	1884	10	US 09 853 386 62	Sequence 62, Appl
59	84.8	4.5	1884	12	US 10-005 169 1	Sequence 1, Appli
60	84.8	4.5	3274	9	US 10 259-864-5	Sequence 5, Appli
61	84.8	4.5	6177	10	US-09-916-145-1	Sequence 1, Appli
62	84.8	4.5	6322	10	US-09-917-800A-1546	Sequence 1546, Ap
63	84.8	4.5	7257	9	US-10-001-486B-1	Sequence 1, Appli
64	83.2	4.4	1606	10	US-09-853-386-55	Sequence 55, Appl
65	83.2	4.4	3802	10	US-09-853-386-46	Sequence 46, Appl
66	83.2	4.4	3802	10	US-09-853-386-78	Sequence 78, Appl
67	83.2	4.4	5115	10	US 09 853 386 57	Sequence 57, Appl
68	81	4.3	2799	10	US-09-935-368 2	Sequence 2, Appli
69	80.8	4.3	1577	10	US-09-797-727-1	Sequence 1, Appli
70	80.8	4.3	2469	10	US-09-853-386-90	Sequence 90, Appl
71	79.2	4.2	3421	10	US 09 853 386 14	Sequence 14, Appl
72	79.2	4.2	3427	9	US-10-171-581-253	Sequence 253, App
73	79.2	4.2	3427	10	US-09-853-386-4	Sequence 4, Appli
74	78.4	4.1	4400	10	US-09-853-386-47	Sequence 47, Appl
75	78.4	4.1	4400	12	US-10-005-169-3	Sequence 3, Appli
76	75.2	4.0	409	10	US-09-960-352-7483	Sequence 7483, Ap
77	74.4	3.9	2154	10	US 09 853 386 24	Sequence 24, Appl
78	74.4	3.9	2247	10	US-09-853-386-29	Sequence 29, Appl
79	74.4	3.9	2353	10	US 09 853 386 23	Sequence 23, Appl
80	74.4	3.9	2458	10	US-09-853-386 3	Sequence 3, Appli
81	74.4	3.9	2907	10	US-09-954-456-524	Sequence 524, App
82	73.4	3.9	1597	9	US 10 153 668 215	Sequence 215, App
83	73	3.8	785	10	US 09 853 386 60	Sequence 60, Appl
84	71.2	3.8	1375	9	US 10 239 804 7	Sequence 7, Appli
85	71.2	3.8	1399	9	US 10 239 804 8	Sequence 8, Appli
86	71.2	3.8	9127	9	US 10 239 804 9	Sequence 9, Appli
87	71.2	3.8	9151	9	US 10 239 804 10	Sequence 10, Appl
88	70.2	3.7	355	9	US 10-081 563 3	Sequence 3, Appli
89	70.2	3.7	42000	9	US-10-081-563-25	Sequence 25, Appl
90	69.6	3.7	2966	10	US-09-853-386-2	Sequence 2, Appli
91	66.6	3.5	1594	10	US 09 964 824A 559	Sequence 559, App
92	65.4	3.4	1431	9	US 10 188 721 2	Sequence 2, Appli

93	65.4	3.4	2218	10	US-09-880-107-3411	Sequence 3411, Ap
94	65.4	3.4	2220	9	US-10-182-242-6	Sequence 6, Appli
95	64.8	3.4	1404	10	US-09-965-703-62	Sequence 62, Appl
96	64.8	3.4	2711	10	US-09-952-559-2	Sequence 2, Appli
97	64.4	3.4	481	10	US-09-864-761-16445	Sequence 16445, A
98	64	3.4	430	9	US-09-918-995-32194	Sequence 32194, A
99	62.2	3.3	549	9	US-09-991-936-1417	Sequence 1417, Ap
100	62	3.3	1759	10	US-09-962-436-300	Sequence 300, App
101	62	3.3	5642	12	US-10-002-600-53	Sequence 53, Appl
102	60.8	3.2	198	10	US-09-965-703-41	Sequence 41, Appl
103	60.8	3.2	1247	10	US-09-965-703-70	Sequence 70, Appl
104	60.8	3.2	1288	10	US-09-965-703-1	Sequence 1, Appli
105	60.8	3.2	1290	10	US-09-965-697-4	Sequence 4, Appli
106	60.8	3.2	1542	10	US-09-965-703-58	Sequence 58, Appl
107	60.8	3.2	2095	12	US-10-142-373-1	Sequence 1, Appli
108	60.4	3.2	987	10	US-09-965-703-21	Sequence 21, Appl
109	60.4	3.2	1398	10	US-09-909-672-1	Sequence 1, Appli
110	60.2	3.2	2511	10	US-09-919-497-36	Sequence 36, Appl
111	60	3.2	392	9	US-10-052-092-4	Sequence 4, Appli
112	60	3.2	1237	9	US-10-052-092-3	Sequence 3, Appli
113	58.8	3.1	463	9	US-09-918-995-15693	Sequence 15693, A
114	58.8	3.1	1740	9	US-10-161-803-26	Sequence 26, Appl
115	57	3.0	2751	10	US-09-728-422-2	Sequence 2, Appli
116	56	3.0	486	9	US-09-918-995-13740	Sequence 13740, A
117	56	3.0	1123	10	US-09-965-703-26	Sequence 26, Appl
118	55.8	2.9	2066	10	US-09-895-840-1	Sequence 1, Appli
119	55.8	2.9	2126	9	US-10-295-370-1	Sequence 1, Appli
120	55.2	2.9	571	9	US-10-066-543-458	Sequence 458, App

ALIGNMENTS

```
RESULT 1
US-10-198-785-1
; Sequence 1, Application US/10198785
; Publication No. US20030022224A1
; GENERAL INFORMATION:
;  APPLICANT: OLYMPUS OPTICAL CO., LTD.
;  TITLE OF INVENTION: Method of detecting binding reaction between protein and test sub
;  FILE REFERENCE: 7EI-02S0538
;  CURRENT APPLICATION NUMBER: US/10/198,785
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: JP/2001 220444
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: JP/2001-221963
;  PRIOR FILING DATE: 2001-07-23
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1740
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (99)..(1688)
US-10-198-785-1

  Query Match             90.5%;  Score 1718;  DB 9;  Length 1740;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1718;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 82

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 142

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 202

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 262

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 322

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 382

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 442

Qy        421 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 CACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAG 502

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 562

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAG 622

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 682

Qy        661 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 AAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 742

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 802

Qy        781 CCACCAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CGACCAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGT 862

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 922

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 982

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAA 1042

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTG 1102

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1162

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGA 1222

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1223 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1282
```

```
Qy       1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGCTCAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1283 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCAC 1342

Qy       1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1343 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1402

Qy       1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATCGCCCT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1403 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATCGCCCT 1462

Qy       1441 GGCTAACCTCCTGATGCTCCTGTCCCAGGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1463 GGCTAACCTCCTGATGCTCCTGTCCCAGGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1522

Qy       1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1523 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1582

Qy       1561 GAATGCCCACAGTGCTTCGGGGGTGCAAGATCCTCCATCACGGGGTCCGAGTGCAGCCCCGC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1583 GAATGCCCACAGTGCTTCGGGGGTGCAAGATCCTCCATCACGGGGTCCGAGTGCAGCCCCGC 1642

Qy       1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTCACGCCTGGCCCT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1643 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTCACGCCTGGCCCT 1702

Qy       1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1718
             |||||||||||||||||||||||||||||||||||||||
Db       1703 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1740


RESULT 2
US-09-815-068A-1
; Sequence 1, Application US/09815068A
; Patent No. US20020072076A1
; GENERAL INFORMATION:
;  APPLICANT: OLYMPUS OPTICAL CO., LTD.
;  TITLE OF INVENTION: Method for testing a substance interacting with hormone receptor
;  FILE REFERENCE: 00S0496P
;  CURRENT APPLICATION NUMBER: US/09/815,068A
;  CURRENT FILING DATE:  2001-03-22
;  PRIOR APPLICATION NUMBER: JP/11 209860
;  PRIOR FILING DATE: 1999-07-23
;  PRIOR APPLICATION NUMBER: JP/2000 163475
;  PRIOR FILING DATE: 2000-05-31
;  PRIOR APPLICATION NUMBER: JP/2000-163476
;  PRIOR FILING DATE: 2000-05-31
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1740
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (99)..(1688)
US-09-815-068A-1

  Query Match           90.5%;  Score 1718;  DB 10;  Length 1740;
  Best Local Similarity  100.0%;  Pred. No. 0:
  Matches 1718;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 82

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 142

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 202

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 262

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 322

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 382

Qy        361 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 ATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 442

Qy        421 CACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGCGCCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 CACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGCGCCAG 502

Qy        481 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTA 562

Qy        541 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTAAAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 CGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTAAAAG 622

Qy        601 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 AAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAA 682

Qy        661 AAACCGGCGCAAGAGCTGCCAGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 AAACCGGCGCAAGAGCTGCCAGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGT 742

Qy        721 GAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GAACTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTCCCGACACACACAAGTGC 802

Qy        781 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCGCGCCCGAGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACCGCGCCCGACT 862

Qy        841 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGC 922

Qy        901 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGAT 982

Qy        961 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAACAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 GATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAACAA 1042

Qy       1021 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGCAGAGCTGTTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 GATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGCAGAGCTGTTG 1102

Qy       1081 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCAT 1162

Qy       1141 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 CTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGA 1222

Qy       1201 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGACTTAAAACTCCAACACAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1223 AATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAA 1282

Qy        1261 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTGCAGTATGTACGCTCTGGTCAC 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1283 AGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACGCTCTGGTCAC 1342

Qy        1321 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1343 AGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGA 1402

Qy        1381 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1403 TGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCT 1462

Qy        1441 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1463 GGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACA 1522

Qy        1501 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1523 TCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCT 1582

Qy        1561 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1583 GAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGCCCGGC 1642

Qy        1621 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1643 AGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGACGCCTGGCCCT 1702

Qy        1681 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1718
               |||||||||||||||||||||||||||||||||||||
Db        1703 GAGGTGAACTGGCCCACAGAGGTCACAAGCTGAAGCGT 1740


RESULT 3
US-09-864-761-7903
; Sequence 7903, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Accmica X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,812
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-10
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7903
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL -   6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 5.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL   5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL - 4.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL   4.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL -   6
US-09-864-761-7903

  Query Match             24.1%;  Score 438;  DB 10;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-126;
  Matches  438,  Conservative    0,  Mismatches    0,  Indels    0;  Gaps    0;

Qy          1 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 CACGAATCTTTGAGAACATTATAATGACCTTTGTGCCTCTTCTTGCAAGGTGTTTTCTCA 76

Qy         61 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GCTGTTATCTCAAGACATGGATATAAAAAACTCACCATCTAGCCTTAATTCTCCTTCCTC 136

Qy        121 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 CTACAACTGCAGTCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTC 196

Qy        181 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTATGTAGACAGCCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAA 256

Qy        241 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 TTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAG 316

Qy        301 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTT 376

Qy        361 ATCACATCTGTATCCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 ATCACATCTGTATCCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACA 436

Qy        421 CACCTTACCTGTAAACAG 438
              ||||||||||||||||||
Db        437 CACCTTACCTGTAAACAG 454


RESULT 4
US-09-893-666A-1
; Sequence 1, Application US/09893666A
; Patent No. US20020129386A1
; GENERAL INFORMATION:
;  APPLICANT: YAMASHITA, ICHIRO
;  TITLE OF INVENTION: High estrogen sensitive medaka fish
```

```
;  FILE REFERENCE: 2102170S-620-7249-0
;  CURRENT APPLICATION NUMBER: US/09/893,666A
;  CURRENT FILING DATE:  2002-02-12
;  PRIOR APPLICATION NUMBER: JP 2000-247729
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 2764
;   TYPE: DNA
;   ORGANISM: Oryzias latipes
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (211)..(1935)
;   OTHER INFORMATION:
US-09-893-666A-1

  Query Match             20.9%;  Score 395.8;  DB 10;  Length 2764;
  Best Local Similarity   62.9%;  Pred. No. 4.7e-113;
  Matches  704;  Conservative    0;  Mismatches  367;  Indels   48;  Gaps    4;

Qy        509 GATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGC 568
              ||  | |  |||||||| || ||||||| ||||| || || || |||||||| || |||
Db        619 GACACGCGTTTCTGCGCCCTGTGCAGCGACTACGCCTCTGGGTACCACTATGGGGTGTGC 678

Qy        569 TCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATT 628
              || ||||| || || |||||||| || || || ||||| || || || ||||| |||||
Db        679 TCTTGTGAGGGCTGCAAGGCCTTCTTCAAGAGGAGCATCCAGGGTCACAATGACTATATG 738

Qy        629 TGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGC 688
              || ||||| || |||||||| || || || | ||| ||| | ||||||||||||||| ||
Db        739 TGCCCAGCGACCAATCAGTGCACTATTCACAGAAATCGGAGGAACAGCTGCCAGGCTTGT 798

Qy        689 CGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG------------------- 729
              || ||| |||||||||||||||||||||||||| ||||  | ||
Db        799 CGTCTTAGGAAGTGTTACGAAGTGGGAATGATGAAAGCGGTGTGCCAAGGACCGCATT 856

Qy        730 -----CTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTCCCGAC 784
                   |  ||| | || | | |  ||    || |||      || | |    ||   |
Db        859 CGCATTTTACGGCGTGACAAACGGCGGACAGGCGTTGGTGATGGAGACAAGGTTGTAAAG 918

Qy        785 GAGCAGCTGCACTGTGCCG---------------GCAAGGCCAAGAGAAGTGCCCGCCAC 829
              |  |||  |||     | |                | ||   ||  || |  || ||
Db        919 GGTCAGGAGCATAAAACGGTGCATTATGATGGAAGGAAACGCAGCAGCACAGGAGGAGGA 978

Qy        830 GCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACC 889
              |     |||  | |    |||  |  ||  |  ||     ||| ||||  |
Db        979 GGAGGAGGAGGAGGAAGACTGTCTGTGACCAGCATACCTCCTGAGCAGGTGCTGCTC 1038

Qy        890 CTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCG------CCAGTGCGCCCTTC 943
              |||||  |||  |   |||||   | ||   |   || |         ||    || | |
Db        1039 CTCCTTCAGGGCCCGAGCCCCCGATACTCTGCTCGCGTCAGAAGTTCAGCCGACCGTAC 1098

Qy        944 ACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATG 1003
              ||||||| | |||||||||     || ||||   |||| |||||||| |||| ||||||
Db        1099 ACCGAGGTCACCATGATGACCCTGCTCACCAGCATGGCAGACAAGGAGCTGGTCCACATG 1158

Qy        1004 ATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGG 1063
               |||  |||||||||||| | || || ||| || ||||   ||||  ||| || |||| |
Db        1159 ATCGCCTGGGCCAAGAAGCTCCCAGGTTTTCTGCAGCTGTCCCTGCACGATCAGGTGCTG 1218

Qy        1064 CTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGAC 1123
               ||  |||||||||  ||| |||||||| | |||||  || || || ||| | || ||  ||
Db        1219 CTGCTGGAGAGCTCCTGGCTGCAGGTGCTCATGATCGGGCTCATTTCCACGCTCCATCCAC 1278

Qy        1124 CACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACACGCATGAGGGGAAATGC 1183
                  ||||| ||||||||||||| | ||| ||  | |||||||| |||||||  | |||
Db        1279 TGTCCCGGGAAGCTCATCTTTGCCACAAGACCTCATCCTGGACAGGAATGAGGGAGACTGC 1338

Qy        1184 GTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAG 1243
               || ||||| ||   ||| ||||| |||||||| ||||| ||| ||||  | || || | |
Db        1339 CTGGAAGGCATGACCGACATCTTCGACATGCTGCTGGCCACTGCTTCCCGCTTCCGTGTG 1398

Qy        1244 TTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGT 1303
                 | |||||| |||   | ||||   ||||  |||| || || ||  | ||||| ||| ||
Db        1399 GTCAAACTCAAACCTGAGGAATTCGTCTGCCTCAAAGCTATTATTTTACTCAACCAGGGT 1458

Qy        1304 ATGTACCCTCTGGTCACAGCGACCCAGGATGC---TGACAGCAGCCGGAAGCTGGCTCAC 1360
                  |   || |   ||| |  |||  |||  |   | ||| ||||  |  | |      |
Db        1459 GCTTTTTCTTTCTGCACCGGCACCATGGAGCCACTTCACAACAGCGGGGGGGTTCAGAGC 1518

Qy        1361 TTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCC 1420
                ||||| || || | || || ||  |  ||       ||      ||  ||  ||  ||
Db        1519 ATGCTGGACACCATCACAGACGCACTCATTCATTACATCAGTCAGTCGGGTTACTTGGCC 1578

Qy        1421 CAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCG 1480
               ||| |||| | |  || | ||| | || ||  |||| || |||||| |||||||   |
Db        1579 CAGGAGCAGGCGAGACGGCAGGCCCAGCTGCTCCTGCTGCTCCCACATCAGGCACATG 1638

Qy        1481 AGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTAT 1540
               || |||| |||||||| || ||   || ||||||||||| ||    ||||  | |||
Db        1639 AGCAACAAAGGCATGGAGCACCTCTACAGCATGAAGTGCAAGAACAAAGTCCCTCTTTAT 1698

Qy        1541 GACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGC 1579
                ||||| || |||||||||||  ||||||||   || | |
Db        1699 GACCTCCTACTGGAGATGCTCGATGCCCACCGCCTGCAC 1737


RESULT 5
US-09-853-033-1
; Sequence 1, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1788
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1788)
US-09-853-033-1

  Query Match             20.6%;  Score 390.8;  DB 10;  Length 1788;
  Best Local Similarity   62.4%;  Pred. No. 1.3e-111;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || |||  || || || || || || || |||||||||
Db        538 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 597

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| |||||||||||||| || |||
Db        598 TGGTGCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 657

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||  || || ||||| || || |||||||| || | ||||||||||||||
Db        658 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 717

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db        718 TGCCGGCTCCGCAAATGCTACGAAGTGGCAATGATGAAAGGTGGGATACGAAAAGACCGA 777

Qy        746 TGTGG------------GTACCGCCTTGTGCGGAGACAGAGAAGTCCCGACGAGCAGCTG 793
                | ||              | | | |      | |||| |     | |  | |  |
Db        778 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 837
```

```
Qy        794 CACTGTGCCCGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCCGAGTGCG--          843
              |  | |     | || ||||  || |  ||    |||  |   ||
Db        838 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 897

Qy        844             GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                          | |    |||       ||||||   |||| ||| | ||    || |
Db        898 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 957

Qy        893 CTGGAGGCTGAGCCGCCCCCATGTGCTGATC--AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||    |      |          || |    |||||||  ||
Db        958 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1017

Qy        950 CCCTCGATCATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTTACACATGATCAGC 1009
              || || ||||||   |  ||||||||  |||| |||| |||| |||| ||||||||||| |
Db       1018 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1077

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1078 TGGCCGAAGAGGGTGCCAGGCTTTGTGGATTGACCCTCCATGATCAGGTCCACCTTCTA 1137

Qy       1070 GACAGCCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |   ||| | || | | ||||| || ||  | |||||||| || || ||||
Db       1138 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1197

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              || |||||  | |||||||| |  |  |  |||||||| |  |||| |||| ||||
Db       1198 GGCAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1257

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| ||||| |||||||| ||||| ||  |  ||  |||| ||    | | ||
Db       1258 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1317

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||      | || | | | ||  |||  | || ||  |||| ||||  |  |||||
Db       1318 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1377

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT--GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  |  |   |||| | |  ||   ||    |   |  |   |  | |||
Db       1378 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1437

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || || ||||| |     |||| ||||||   ||| |  ||   ||||||
Db       1438 GACAAGATCACAGACATTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1497

Qy       1427 CAATCCATGGCGCCTGGCTAACTCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| ||||||| |||||||   |||||||
Db       1498 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1557

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCACTGTATGACCTG 1546
              || |||||||| ||||||   || |||||||||| || ||||| ||  | |||||||||
Db       1558 AAAGGCATGGAGCATCTCTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1617

Qy       1547 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCCCCA 1586
              |||||||||||||||| | ||||||   || |  | |  ||
Db       1618 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCCCCA 1657


RESULT 6
US-09-281-674-9
; Sequence 9, Application US/09281674
; Patent No. US20020077307A1
;   GENERAL INFORMATION:
;        APPLICANT: Gossen, Manfred
;                   Bujard, Hermann
;                   Salfeld, Jochen
;                   Voss, Jeffrey
;        TITLE OF INVENTION: Methods for Regulating Gene Expression
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Lahive & Cockfield
;             STREET: 60 State Street, Suite 510
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: ASCII text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/281,674
;             FILING DATE: 30-Mar-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/479,306
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/260,452
;             FILING DATE: 14-JUN-1994
;             APPLICATION NUMBER: 08/076,427
;             FILING DATE: 14-JUN-1993
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Giulio A. DeConti, Jr.
;             REGISTRATION NUMBER: 31,503
;             REFERENCE/DOCKET NUMBER: BBI-013CP3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 227-7400
;             TELEFAX: (617) 227-5941
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4963 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        ORIGINAL SOURCE:
;             ORGANISM: Human cytomegalovirus
;        IMMEDIATE SOURCE:
;             CLONE: pUHD BGR4
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-281-674-9

  Query Match             20.6%;  Score 390.8;  DB 10;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 2.5e-111;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   89;  Gaps    4;

Qy         506 AGGCATGCTCACTTCTGCGCTGTCTGCAGCGATTATGCATCGGGATATCACTATGGAGTC 565
               | |||  ||| || || || || || ||||  || || || || || || || |||||||||
Db        1003 AAGGACACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy         566 TGGTCGTGTGAAGCATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
               ||||| ||||| || || |||||||| || || ||||| ||||||||||||| || |||
Db        1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy         626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
               || |||||||| || || |||| || || |||||||| || | ||||||||||||||
Db        1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy         686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGGAGAGAGAGA 745
               |||| || || || || || |||||||||||||||| |||| ||||    || | |||  ||
Db        1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy         746 TGTGG             GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
                 | ||            | | | |      | |||| |     | |  | | |
Db        1243 AGAGGAGGCAGAATGTTGAAACACAAGCGCCACAGAGACATGATGGGGAGGGCAGGGGTGAA 1302

Qy         794 CACTGTGCCCGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 843
               |  | |     | || ||||  || |  ||    |||  |   ||
Db        1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy         844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                         | |    |||       ||||||   |||| ||| | ||    || |
Db        1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1422

Qy         893 CTGGAGGCTGAGCCGCCCCCATGTGCTGATC--AGCCGCCCCAGTGCGCCCTTCACCGAG 949
```

```
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCAGTATGATCCTACCACAGCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||    | ||||||||  |||| |||| |||| |||| |||||||||| |
Db       1483 GCTTCGATCATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||   || || || |  || |
Db       1543 TGGGCGAAGAGGGTGCCAGCCTTTGTGGATTGACCCTCCATGATCAGGTCACACTTCTA 1602

Qy       1070 GAGAGCTGTTGCATGCAGGTGTTAATGATGCCGCTGATGCCGCTCAATTGACCACCCC 1129
              ||  |   ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTGCTGGGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAGGGGAAATGCGTAGAA 1189
              |  ||||  | ||||||||  |  |  |  |||||||| |  |||| |||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACCTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTCCAGAGCTTAAAA 1249
              || ||  |||| ||||| |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCATGATCCTCCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||| | || ||  |||| |||||  |  |||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCATCCCGAACCTGGCTCACTTGCTG 1366
              |  |  | |   |||| | |  ||   |     |  |     |  | | ||| |||
Db       1843 ACATTTCTGTCCAGCACCTGAAGTCTCTGGAAGAAGAAGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATCTTTGGTTTGGGTCATTGCCAAGAGCGCATCTCCTCCAGTAC 1426
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGTCCCTGTCCCACGTCAGGCATGGAGTAAC 1486
              ||   | ||| |||||  | |||||  |  |||| ||||||  |||||||    |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || |||||||||||| || ||||| ||  |  |||||||||
Db       2023 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGGGGGGTGCA 1586
              |||||||||||||||  | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTGGATGCCCACCGCCTATCATGCCCCA 2122


RESULT 7
US-09-892-227-9
; Sequence 9, Application US/09892227
; Patent No. US20020152487A1
;   GENERAL INFORMATION:
;        APPLICANT: Gossen, Manfred
;                   Bujard, Hermann
;                   Salfeld, Jochen
;                   Voss, Jeffrey
;        TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
;                            Transcriptiona
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Lahive & Cockfield
;             STREET: 60 State Street, Suite 510
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: ASCII text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/892,227
;             FILING DATE: 25-Jun-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/163,269
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/076,327
;             FILING DATE: 14-JUN-1993
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Giulio A. DeConti, Jr.
;             REGISTRATION NUMBER: 31,503
;             REFERENCE/DOCKET NUMBER: BBI-013CP2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 227-7400
;             TELEFAX: (617) 227-5941
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4963 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        ORIGINAL SOURCE:
;             ORGANISM: Human cytomegalovirus
;        IMMEDIATE SOURCE:
;             CLONE: pUHD BGR4
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-892-227-9

  Query Match             28.6%;  Score 700.8;  DB 10;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 2.5e-111;
  Matches  699;  Conservative    0;  Mismatches  482;  Indels   39;  Gaps    4;

Qy        505 AGGGATGCTCACTTCTGGGCTGTCTGCAGGGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db       1004 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACTATTATGGAGTC 1062

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATGGATAAAAACCGGCGCAAGAGCGCCAGGCC 685
              || |||||| || || |||| || || |||||||||| || | ||||||||||||||
Db       1123 ATGTGTCCAGCCACTAACTCAGTGCACTATGGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy        686 TGCCCACTTCCGAAGTCGTTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        746 TCTGG-----------GTACGGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
               | ||           | | | |    | |||| |     | |  | |  |
Db       1243 AGAGGAGCCAGAATCTTGAAACACAAGCGCCAGAGAGATGATGGGGACGGCAGGGTTGAA 1302

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 843
                | | |    | || ||||  || |  ||    |||  |   ||
Db       1303 GTGGGCTGTGCTGGAGACATCAGAGCTGCCAACCTTGGGCAAGCCGCTCATGATCAAA 1362

Qy        844 ----------GCACCTGCTGCTGGACGCCCTGACGCCCGACGAGCTAGTGCTCACCCTC 892
                       | |    |||      ||||||   |||| ||| | ||  |  || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCGGACCAGATGGTCAGCGCCTTG 1422

Qy        893 CTGGAGGCTGAGCCGCCCCATGGCGTATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| ||||||| |||    |     |         || |     |||||||  ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCAGTATGATCCTACCACAGCCTTCAGTGAA 1482

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||    | ||||||||  |||| |||| |||| |||| |||||||||| |
Db       1483 GCTTCGATCATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || ||||||||||  | | |||    || || || | ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGCCTTTGTGGATTTGACCCTCCATGATCAGGTCACACTTCTA 1602
```

```
Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
             ||  |    ||| | ||| |  | ||||| || |   | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1662

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  |||||  | |||||||| |  |  |  ||| |||| |  |||| ||||| |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| |||| |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTGGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||  | || ||  |||| ||||  |  ||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  | |    |||| | |  ||   ||    |     |  |    |  | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |     |||| |||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1427 CAATCCATGCGGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || ||||||||||||| || ||||| ||  | |||||||||
Db       2023 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 8
US-10-052-092-6
; Sequence 6, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Bopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102DS2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;  LENGTH: 2092
;  TYPE: DNA
;  ORGANISM: Human
US-10-052-092-6

  Query Match             29.5%;  Score 489.2;  DB 9;  Length 2092;
  Best Local Similarity   62.3%;  Pred. No. 4.7e-111;
  Matches  698;  Conservative    0;  Mismatches  483;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || || || || || ||||  || || || || || || || |||||||||
Db        840 AAGGAGACTCGCTACTGTGCCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 899

Qy        566 TGGTCGTGTGGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| |||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        900 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 959

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db        960 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1019

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | ||  ||
Db       1020 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1079

Qy        746 TGTGG- --------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||            | | | |      | |||| |      | |  | |  |
Db       1080 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1139

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG - - - - 843
              |  | |      | || ||||  || |  ||    |||  |   ||
Db       1140 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1199

Qy        844 ---- ------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                   | |   |||        ||||||   |||| ||| | ||     || |
Db       1190 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1249

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||   |     |          || |     |||||||  ||
Db       1250 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1309

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  ||||||||  |||| |||| |||| |||| ||||||||||| |
Db       1310 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1369

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              |||| ||| | | || |||||||||||  | | |||    || || || |  ||  |
Db       1370 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1429

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1430 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1489

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  |||||  | |||||||| |  |  |  |||||||| |  |||| ||||| |||||
Db       1490 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1549

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| |||| |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       1550 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1609

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||  | || ||  |||| |||||  |  ||||
Db       1610 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1669

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  | |    |||| | |  ||   ||    |     |  |    |  | |||
Db       1670 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1729

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |     |||| |||||   ||| |  ||   ||||||
Db       1730 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1789

Qy       1427 CAATCCATGCGGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1790 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1849

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || ||||||||||||| || ||||| ||  | |||||||||
Db       1850 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1909

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||| | ||||||   || |  | |  ||
Db       1910 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 1949


RESULT 9
US-09-952-680A-9
; Sequence 9, Application US/09952680A
```

```
; Publication No. US20030087239A1
; GENERAL INFORMATION:
;  APPLICANT: Stanton, Marty
;  APPLICANT:  Epstein, David
;  APPLICANT:  Hamaguchi, No. US20030087239Aluko
;  TITLE OF INVENTION: Target Activated Biosensor and Methods of Using Same
;  FILE REFERENCE: 23239-501
;  CURRENT APPLICATION NUMBER: US/09/952,680A
;  CURRENT FILING DATE:  2001-09-13
;  PRIOR APPLICATION NUMBER: 60/232,454
;  PRIOR FILING DATE: 2000-09-13
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 2092
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-952-680A-9

  Query Match             20.5%;  Score 389.2;  DB 9;  Length 2092;
  Best Local Similarity   62.3%;  Pred. No. 4.7e-111;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        830 AAGGACACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 889

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        890 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 949

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || ||||||||| || | ||||||||||||||||
Db        950 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1009

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGGACACAGACA 745
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db       1010 TGCCGGCTCCGCAAATGCTACGAAGTGGCAATGATGAAAGTGTGGGATACGAAAACACCGA 1069

Qy        746 TGTGG-----------GTACGGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||           | | | |      | |||| |     | | | | |
Db       1070 AGAGGAGGCAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGCAGGGCAGGGTGAA 1129

Qy        794 CACTGTGCCGGCAAGGCCAAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 843
                | | |     | || |||| || | ||    ||| | ||
Db       1130 GTGGGGTCTGCTGGAGACATCAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1189

Qy        844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTACTGCTCACCCTC 893
                     | |    |||      ||||||   |||| ||| | ||     || |
Db       1190 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1249

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCGCGCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||    |   |          || |     ||||||| ||
Db       1250 TTGGATGCTGAGCCCCCCATACTCTATTCCGACTATGATCCTACCAGACCCTTCAGTGAA 1309

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAAC 1009
              || || ||||||   |  |||||||| |||| |||| |||| |||| |||||||||| |
Db       1310 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1369

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTC 1069
              ||||| ||| | | || |||||||||||  | | |||    || || || |  || |
Db       1370 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1429

Qy       1070 GAGAGCTGTTGGATGCAGGTGTTAATGATGGGGCTGATCTGGCCCTCAATTCACCAGCCC 1129
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1430 GAATGTGCCTGGCTACAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGCAGCACCCA 1489

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTACAA 1189
              |  ||||| | |||||||| |  | |  | |||||||| |  |||| ||||| |||||
Db       1490 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTACAG 1549

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| ||||| |||||||| |||| ||  | ||  |||| ||   | | ||

Db       1550 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1609

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||| | || ||  |||| |||||  |  |||||
Db       1610 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1669

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  | | |    |||| | |  ||   ||  |       |  |  |    |  | |||
Db       1670 ACATTTCTGTCCAGCACCCTGAAGTCTCTGCAACAGAAGGACCATATCCACCCAGTGCTG 1729

Qy       1367 AACCCCGTGACCCATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
                ||    | || ||  ||||| |     |||| ||||||   ||| |  ||  ||||||
Db       1730 GACAAGATCACAGACACTTTCATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1789

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATCGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1790 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1849

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| ||||||  || |||||||||||| || |||| ||  | |||||||||
Db       1850 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1909

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||| | ||||||   || |  | |  ||
Db       1910 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 1949


RESULT 10
US-10-096-710-2
; Sequence 2, Application US/10096710
; Patent No. US20020164581A1
; GENERAL INFORMATION:
;  APPLICANT: Chang, Chawnshang
;  TITLE OF INVENTION: METHODS FOR IDENTIFYING COMPOUNDS THAT
;  TITLE OF INVENTION:  DO NOT ATTENUATE THE PROTECTIVE EFFECTS OF ESTROGEN
;  FILE REFERENCE: 21108.0010U2
;  CURRENT APPLICATION NUMBER: US/10/096,710
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/275,229
;  PRIOR FILING DATE: 2001-03-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 6450
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:/note =
;   OTHER INFORMATION: synthetic construct
US-10-096-710-2

  Query Match             20.5%;  Score 389.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 9.2e-111;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        898 AAGGACACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || |||| || || ||||||||| || | ||||||||||||||||
Db       1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGGACACAGACA 745
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db       1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGTGTGGGATACGAAAACACCGA 1137

Qy        746 TGTGG-----------GTACGGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||           | | | |      | |||| |     | | | | |
```

```
Db        1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1197

Qy         794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG------------ 843
               |  | |      | || |||| || |  ||    ||| |   ||
Db        1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1257

Qy         844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                      | |    |||       ||||||   |||| ||| | ||     || |
Db        1258 CGCTCTAAGAAGAACAGCCTGGCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCTTG 1317

Qy         893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
               |||| ||||||||| |||    |       |          || |     |||||||  ||
Db        1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy         950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
               || || ||||||   |  |||||||  |||| |||| |||| |||| ||||||||| |
Db        1378 GCTTCGATGATGGGCTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy        1010 TGGGCCAAGAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
               ||||| |||| | | || ||||||||||||  | | |||    || || || |  ||  |
Db        1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy        1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
               ||  |    ||| | ||| |  | ||||| || ||  | ||||||| || || |||||
Db        1498 GAATGTGCCTGGCTAGACATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1557

Qy        1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
               |  ||||| | |||||||  |  |  |  |||||||| |  |||| |||| |||||
Db        1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy        1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
               || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db        1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677

Qy        1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
               || ||    | || | | | ||  |||| | || ||  |||| |||||  |  |||||
Db        1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1737

Qy        1310 CCTCTGGTCACAGCGACCCAGGATGCT--GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
               |  |  | |    |||| | |  ||   ||    |     |  |    |  | |||
Db        1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGACAAGGACCATATCCACCGAGTCCTG 1797

Qy        1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
               ||    | || ||  .|||| |      |||| |||||   ||| |  ||   |||||
Db        1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy        1427 CAATCCATGGCGCTGGCTAACCTCCTGATGTCCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
               ||   | |||  |||| | |||||  | |||| ||||||   |||||||    |||||||
Db        1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy        1487 AAGGGCATGGAACATCTGCTCAACATCAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
               || |||||||| |||||  || ||||||||||||| || |||| ||  | |||||||||
Db        1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy        1547 CTGCTGGAGATGCTGGAATGCCACGTGCTTCGCGGGTGCA 1586
               |||||||||||||||| | ||||||   || |  | |  ||
Db        1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 11
US-10-081-563-1
; Sequence 1, Application US/10081563
; Publication No. US20020187495A1
; GENERAL INFORMATION:
;  APPLICANT: Wake Forest University
;  APPLICANT:  Herrington, David M.
;  APPLICANT:  Howard, Timothy D.
;  APPLICANT:  Hawkins, Gregory A.
;  APPLICANT:  Moyers, Deborah A.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS OF ESTROGEN RECEPTOR ALPHA ASSOCIATED WITH
;  TITLE OF INVENTION:  FAVORABLE HDL CHOLESTEROL RESPONSE TO HORMONE REPLACEMENT THERAP
;  FILE REFERENCE: 9151.15
;  CURRENT APPLICATION NUMBER: US/10/081,563
;  CURRENT FILING DATE:  2002-02-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;  LENGTH: 6450
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (361)..(2148)
;  OTHER INFORMATION:
;  PUBLICATION INFORMATION:
;  DATABASE ACCESSION NUMBER: X03635
;  DATABASE ENTRY DATE: 1993-09-12
;  RELEVANT RESIDUES: (1)..(6450)
US-10-081-563-1

  Query Match             20.5%;  Score 389.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 9.2e-111;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy         506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
               | |||  ||| || ||| || || ||||  || || || || || || || || |||||||||
Db         898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy         566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
               ||||| |||| || || |||||||| || || |||| |||||||||||||| || |||
Db         958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy         626 ATTIGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCCAAGAGCTGCCAGGCC 685
               || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db        1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy         686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
               ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db        1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1137

Qy         746 TGTGG----------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
               | ||           | | | |    | |||| |     | |  | |  |
Db        1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1197

Qy         794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG------------ 843
               |  | |      | || |||| || |  ||    ||| |   ||
Db        1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1257

Qy         844 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                      | |    |||       ||||||   |||| ||| | ||     || |
Db        1258 CGCTCTAAGAAGAACAGCCTGGCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCTTG 1317

Qy         893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
               |||| ||||||||| |||    |       |          || |     |||||||  ||
Db        1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy         950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
               || || ||||||   |  |||||||  |||| |||| |||| |||| |||||||||  |
Db        1378 GCTTCGATGATGGGCTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy        1010 TGGGCCAAGAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
               ||||| |||| | | || ||||||||||||  | | |||    || || || |  ||  |
Db        1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy        1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
               ||  |    ||| | ||| |  | ||||| || ||  | ||||||| || || |||||
Db        1498 GAATGTGCCTGGCTAGACATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1557

Qy        1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
               |  ||||| | |||||||  |  |  |  |||||||| |  |||| |||| |||||
Db        1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy        1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
               || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db        1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677
```

```
Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
             || ||     | || | | | | ||  |||| | || ||  |||| ||||  |  |||||
Db       1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1737

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              | | | |    |||| ) |  ||   ||   |    |  | |    | |||
Db       1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  ||||| |     |||| ||||||   ||| |  ||   ||||||
Db       1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| ||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || |||||||| |||||| || ||||||||||| || ||||| ||  | |||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy       1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||| | ||||||   || |  | |  ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 12
US-10-052-092-1
; Sequence 1, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 6450
;   TYPE: DNA
;   ORGANISM: Human
US-10-052-092-1

  Query Match             20.5%;  Score 389.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 9.2e-111;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || ||||  || || || || || || || ||||||||||
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAAIGATTAT 625
              ||||| |||| || || |||||||| || || |||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACCACTAT 1017

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGGC 685
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db        1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGGC 1077

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || ||||||||||||||| ||||  ||||    || | ||  ||
Db        1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGCATACGAAAACACCCA 1137

Qy        746 TGTGG----------CTACCGCCTTGTGCCGAGACAGAGAAGTGCCGACCAGCAGCTC 793
               | ||          | | | |     | |||| |     | |  | |  |
Db        1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGCTGAA 1197

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG ------ 843
                | | |     | || |||  || |  ||    |||  |   ||
Db        1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCGCTCATGATCAAA 1257

Qy        844 -----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                        | |    |||       ||||||   |||| ||| | ||     || |
Db        1258 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1317

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| |||||||| |||    |        |           || |     |||||||  ||
Db        1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   |  |||||||| |||| |||| |||| |||| ||||||||||| |
Db        1378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy        1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
               ||||| |||| | | || ||||||||||||  | | |||    || || || |  ||  |
Db        1438 TGGGGCAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy        1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGGGCTCAATTGACCACCCC 1129
               || |    ||| | || |  | ||||| || ||  | |||||||| || || |||||
Db        1498 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTGCTCTGGCGCTCCATGGAGCACCCA 1557

Qy        1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
               |  ||||  | ||||||||  | |  |  |||||||| |  |||| |||| |||||
Db        1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy        1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
               || ||  |||| |||||  |||||||| |||| ||  | ||  |||| ||   | | ||
Db        1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677

Qy        1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
               || ||     | || | | | | ||  |||| | || ||  |||| ||||  |  |||||
Db        1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1737

Qy        1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
               | | | |    |||| ) |  ||   ||   |    |  | |    | |||
Db        1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy        1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
               ||    | || ||  ||||| |     |||| ||||||   ||| |  ||   ||||||
Db        1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy        1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
               ||   |  ||| ||||  | |||||  |  |||| |||||| |||||||   |||||||
Db        1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy        1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
               || |||||||| |||||| || ||||||||||| || ||||| ||  | |||||||||
Db        1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy        1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1586
               |||||||||||||||| | ||||||   || |  | |  ||
Db        1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 13
US-10-052-092-7
; Sequence 7, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
```

```
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 6450
;   TYPE: DNA
;   ORGANISM: Human
US-10-052-092-7

  Query Match             20.5%;  Score 389.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 9.2e-111;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || |||  || || || || || || || |||||||||
Db       2898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 2957

Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db       2958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGCACATAACGACTAT 3017

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || ||||| || || ||||||||| || | |||||||||||||||
Db       3018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 3077

Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || |||||||||||||| |||| ||||    || | ||| ||
Db       3078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 3137

Qy        746 TGTGG------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||      | | | |     | |||| |     | |  | |  |
Db       3138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 3197

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG    --      843
              | | |     | || |||  || |  ||   |||  |   ||
Db       3198 GTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTGGCCAAGCCCGCTCATGATCAAA 3257

Qy        844 -----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 892
                  | |    |||       ||||||   |||| ||| | ||     || |
Db       3258 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 3317

Qy        893 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 949
              |||| ||||||||| |||    |       |          || |     |||||||  ||
Db       3318 TTGGATGCTGAGCCCCCCATACTCTATTCGGAGTATGATCCTACCAGACCCTTCAGTGAA 3377

Qy        950 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 1009
              || || ||||||   | ||||||||  |||| |||| |||| |||| ||||||||||| |
Db       3378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 3437

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069
              ||||| |||| | | || |||||||||||  | | |||    || || || |  ||  |
Db       3438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 3497

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129
              ||  |    ||| | || |  | ||||| || ||  | |||||||| || || |||||
Db       3498 GAAPGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 3557

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 1189
              |  |||||  | |||||||  | |  |  |||||||  |  |||| |||||  |||||
Db       3558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 3617

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249
              || ||  |||| |||||  |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       3618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 3677

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309
              || ||     | || | | | ||  |||| | || ||  |||| |||||  |  |||||
Db       3678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 3737

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1366
              |  |  | |    |||| | |  ||    |     | |     |  | |||
Db       3738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTC 3797

 .       .
```

```
Qy       1367 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||  |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy       1487 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
              || ||||||| |||||  || |||||||||||| || ||||  ||  | ||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy       1547 CTGCTGGAGATGCTGGAATGCCCACGTGCTTCGCGGGTGCA 1586
              |||||||||||||||  | ||||||   || |  | |  ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 14
US-10-052-092-20
; Sequence 20, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapo
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
;   LENGTH: 6610
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Vector pC1-n GLI-HEGO
US-10-052-092-20

  Query Match             20.5%;  Score 389.2;  DB 9;  Length 6610;
  Best Local Similarity   62.3%;  Pred. No. 9.4e-111;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        506 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
              | |||  ||| || ||| || || |||  || || || || || || || |||||||||
Db       2422 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 2481

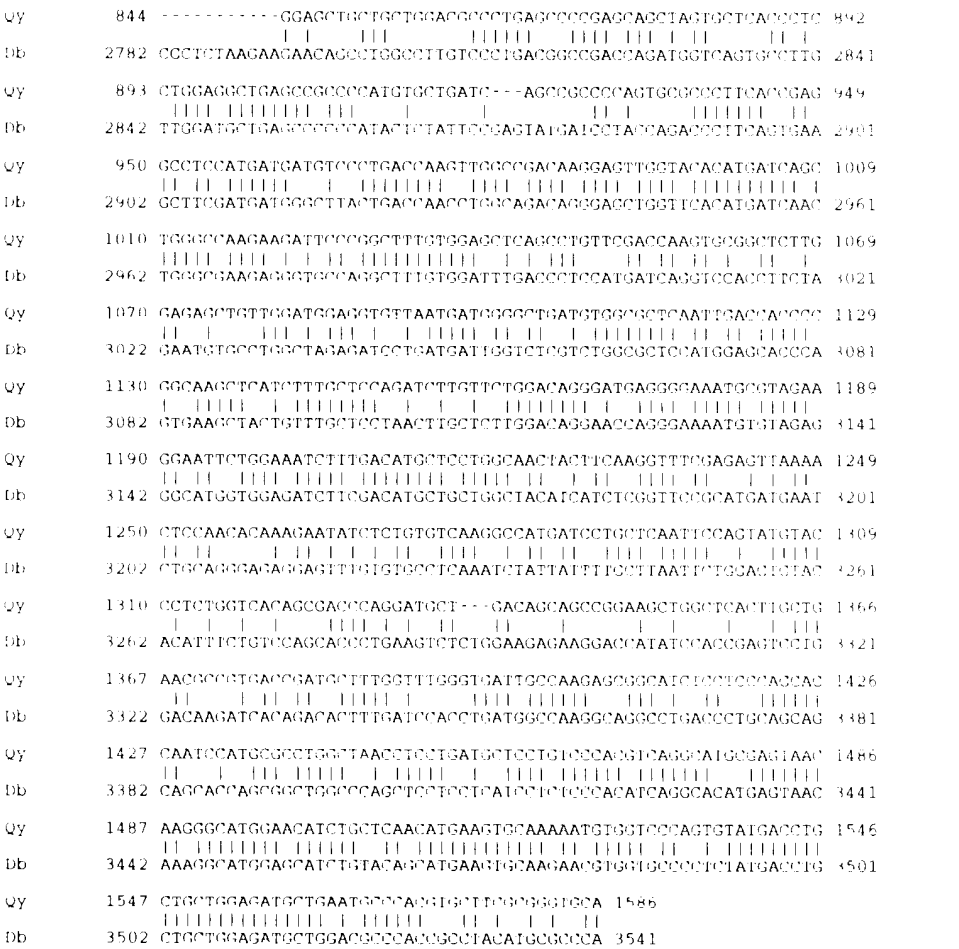
Qy        566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db       2482 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 2541

Qy        626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
              || |||||||| || || ||||| || || ||||||||| || | |||||||||||||||
Db       2542 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 2601

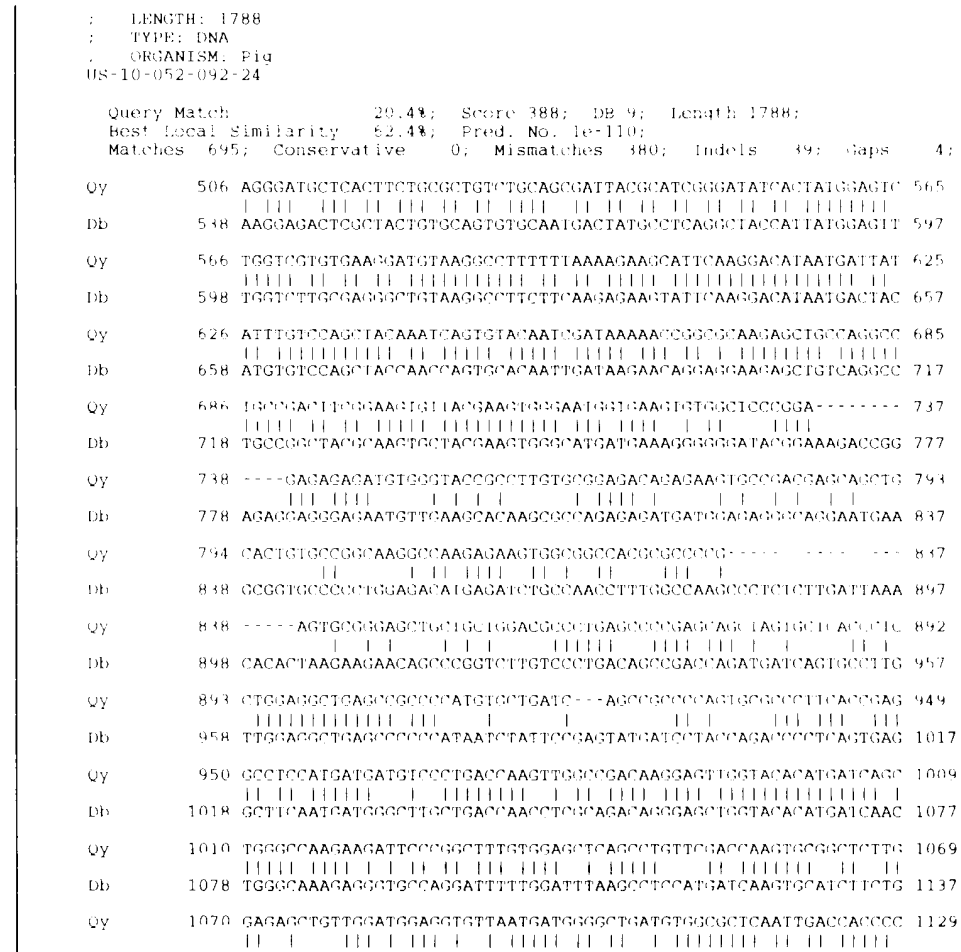
Qy        686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 745
              ||||| || || || || |||||||||||||| |||| ||||    || | ||| ||
Db       2602 TGCCCGCTCCGCCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAACACCCA 2661

Qy        746 TGTGG------------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 793
              | ||                | | | |     | |||| |     | |  | |  |
Db       2662 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 2721

Qy        794 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG    --      843
              | | |     | || |||  || |  ||   |||  |   ||
Db       2722 GTGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 2781
```



RESULT 15
US-10-052-092-24
; Sequence 24, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24



; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Pig
US-10-052-092-24

Query Match 20.4%; Score 388; DB 9; Length 1788;
Best Local Similarity 62.4%; Pred. No. 1e-110;
Matches 695; Conservative 0; Mismatches 380; Indels 89; Gaps 4;

```
Qy        1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486
               ||   |  ||| || ||  | |||||  |  ||||||| ||| |||||||  |||||||
Db        1498 CAGCACCGGCGTCTCGCGCAGCTCCTCCTCATCCTGTCTCACTTCAGGCACATGAGTAAC 1557

Qy        1487 AAGGGCATGGAACATCTGTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1546
               || ||||||||  ||||||  ||||||||||||||| || |||| ||  | |||||||
Db        1558 AAAGGCATGGAGCATCTGTACAACATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1617

Qy        1547 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCG 1580
               ||||||||||||||| | ||||||  || | ||
Db        1618 CTGCTGGAGATGCTGGACGCCCACCGCCTGCACG 1651


RESULT 16
US 10 198 846 11305
; Sequence 11305, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI 049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002 07 18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001 07 18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11305
;   LENGTH: 6608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 6608
;   OTHER INFORMATION: n   A,T,C or G
US 10 198 846 11305

  Query Match             20.3%;  Score 485.2;  DB 9;  Length 6608;
  Best Local Similarity   62.1%;  Pred. No. 1.7e-109;
  Matches  698;  Conservative    0;  Mismatches  483;  Indels   43;  Gaps    4;

Qy         506 AGGGATGCTCACTTCTGCGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 565
               | |||  ||| || ||  || || |||  || || || || || || ||||||||
Db         898 AAGGAGACTCGCTACTGTGCCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy         566 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 625
               ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db         958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy         626 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 685
               || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db        1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy         686 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGGAGACAGAGA 745
               ||||| || || || || ||||||||||||||| |||| ||||    || | ||| ||
Db        1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1137

Qy         746 TGTGGGTACCCGCCTTGTG - - - - - -CGGAGACAGAGAAGTGCCGACGAGCA 789
               | ||     | | ||                ||    | | | | | || | |
Db        1138 AGAGGAGCGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGTACTTGCTACCCGAA 1197

Qy         790 GCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCT 849
               |    | |    | || |||  || |  ||    ||| |   ||    | |
Db        1198 TTCCGATCCATCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGAT 1257

Qy         850 ---------- GCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCAC 888
                          |||      |||||   |||| ||| | | ||     |
Db        1258 CAAACGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGC 1317

Qy         889 CCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCAC 945
               | |  |||| |||||||| |||    |       |          || |     |||||||
Db        1318 CTTGTTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAG 1377

Qy         946 CGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGAT 1005
               || || || ||||||   | ||||||||  |||| |||| |||| |||| |||||||||
Db        1378 TGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGAGCTGGTTCACATGAT 1437

Qy        1006 CAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCT 1065
               || |||||| |||| | | || ||||||||||  | | |||    || || || |  ||
Db        1438 CAACTGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCT 1497

Qy        1066 CTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCA 1125
                 | ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || ||
Db        1498 TCTAGAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCA 1557

Qy        1126 CCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGT 1185
               ||| |  |||||  | |||||||  |  |  |  |||||||| |  |||| ||||| ||
Db        1558 CCCAGTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGT 1617

Qy        1186 AGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTT 1245
               ||| || ||  |||| |||| ||||||| |||| || | || |||| ||   | ||
Db        1618 AGAGGGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCGGCATGAT 1677

Qy        1246 AAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTAT 1305
               || || ||      | || | | | ||  |||  | || ||  |||| |||| |  |
Db        1678 GAATCTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCIATTATTTTGCTTAATTCTGGAGT 1737

Qy        1306 GTACCCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTT 1362
               |||| |  | |  |   |||| | |  ||   ||    |      | |  |  |   |
Db        1738 GTACACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGT 1797

Qy        1363 GCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCA 1422
               ||| ||    | || ||  |||| |    |||| ||||||   ||| |  ||   ||
Db        1798 CCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCA 1857

Qy        1423 GCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
               ||||||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||
Db        1858 GCAGCAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAG 1917

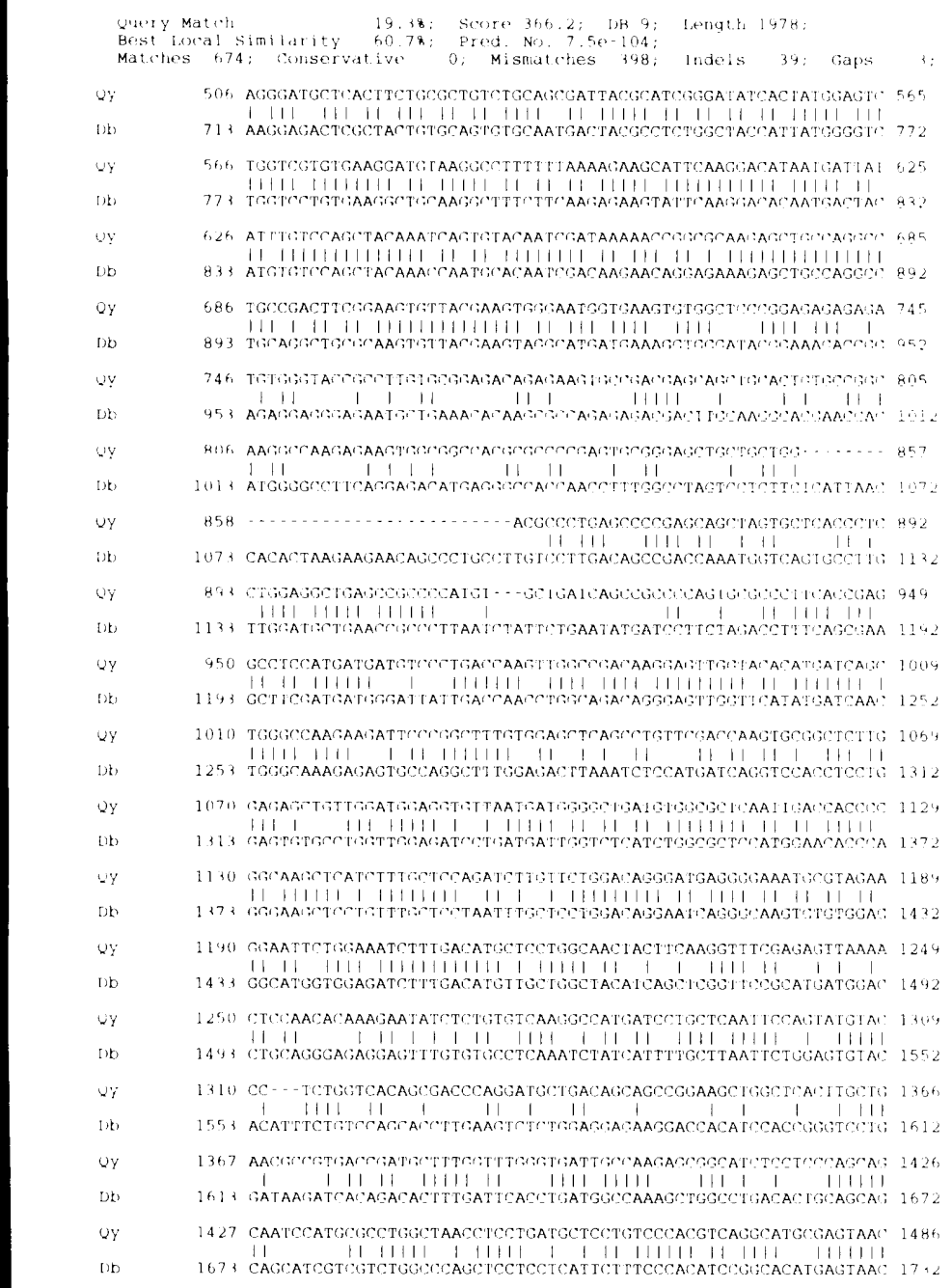
Qy        1483 TAACAAGGGCATGGAACATCTGTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGA 1542
               ||||| ||||||| |||||  || |||||||||| || ||||||||  | |||||  |
Db        1918 TAACAAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGA 1977

Qy        1543 CCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCA 1586
               |||||||||||||||||| | ||||||   || |  | |  ||
Db        1978 CCTGCTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2021


RESULT 17
US 10-052-092-23
; Sequence 23, Application US/10052092
; Publication No. US20040027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapy
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002 01 18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001 01 19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001 07 09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
;   LENGTH: 1978
;   TYPE: DNA
;   ORGANISM: Hamster
```

US-10-052-092-23

Query Match 19.3%; Score 366.2; DB 9; Length 1978;
Best Local Similarity 60.7%; Pred. No. 7.5e-104;
Matches 674; Conservative 0; Mismatches 398; Indels 39; Gaps 4;



RESULT 18
US-10-052-092-5
; Sequence 5, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2106
; TYPE: DNA
; ORGANISM: Human
US-10-052-092-5

Query Match 17.2%; Score 326.6; DB 9; Length 2106;
Best Local Similarity 60.8%; Pred. No. 2.1e-91;
Matches 624; Conservative 0; Mismatches 364; Indels 39; Gaps 4;

```
Db       1018 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1077

Qy       1010 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 1069

Db       1078 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1137

Qy       1070 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 1129

Db       1138 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1197

Qy       1130 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCCTAGAA 1189

Db       1198 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1257

Qy       1190 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1249

Db       1258 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCCGTCCGCCATGATGAAT 1317

Qy       1250 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1309

Db       1318 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1377

Qy       1310 CCTCTGGTCACAGCGACCCAGGATGCT-- GACA3CAGCCGGAAGCTGGCTCACTTGCTG 1366

Db       1378 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1437

Qy       1367 AACGCCGTGACCGATGCTTTGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1426

Db       1438 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1497

Qy       1427 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1486

Db       1498 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGGAAC 1557

Qy       1487 AAGGGCA 1493

Db       1558 CAGGGAA 1564


RESULT 19
US-10-259-864-3
; Sequence 3, Application US/10259864
; Publication No. US20030077645A1
; GENERAL INFORMATION:
;  APPLICANT: Bager, Gordon L.
;  TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfamil
;  TITLE OF INVENTION:  Receptor Ligands, and Methods and Kits For Detecting and Charact
;  TITLE OF INVENTION:  Ligands
;  FILE REFERENCE: 24799
;  CURRENT APPLICATION NUMBER: US/10/259,864
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: US 60/425,178
;  PRIOR FILING DATE: 2001-09-28
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 3300
;   TYPE: DNA
;   ORGANISM: Rat/human Chimera
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(3300)
;   OTHER INFORMATION:
US-10-259-864-3

  Query Match             16.2%;  Score 306.8;  DB 9;  Length 3300;
  Best Local Similarity   57.9%;  Pred. No. 4.4e-85;
  Matches  635;  Conservative    0;  Mismatches  432;  Indels   29;  Gaps    4;

Qy        519 TCTGCGCTGTCTGCAGCGATTACGCATCGGCATATCACTATGGAGTCTGGTCGTGTGAAG 578

Db       2075 TCTGCCTGGTGTGCTCCGATGAAGCTTCAGGATGTCATTACGGGGTGCTGACATGTGGAA 2134

Qy        579 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 638

Db       2135 GCTGCAAAGTATTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 2194

Qy        639 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 698

Db       2195 GAAACGATTGCATCATTGATAAAATTCGAAGGAAAAACTGCCCAGCATGCCGCTATCGGA 2254

Qy        699 AGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCC 758

Db       2255 AATGTCTTCAGGCTGGAATGAACCTTGAAGCTCGAAAAACAAAGAAAAAAATCAAAGGGA 2314

Qy        759 TTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAA 818

Db       2315 TTCAGC-AAGCCACTGCAGGAGTCTCACAAGACACTTCGGAAAATCCTAACAAAACAATA 2373

Qy        819 GTGGCGGCCACGCCCCCCGAGTGCG------ ---GGAGCTGCTGCTGGACGCCCTGAGC 868

Db       2374 GTTCCTGCAGCATTACCACAGCTCACCCCTACCTTGGTGTCACTGCTGGAGGTGATTGAA 2433

Qy        869 CCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGC 928

Db       2434 CCCGAGGTGTTGTATGCAGGATATGATAGCTCTGTTTCGCGAGGTGAGCCCCCCATACTC 2493

Qy        929 ---------------CCCAGTGCGGCCTTCACCGAGGCCTGCATGATGATGTCCCTGACC 973

Db       2494 TATTCCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACC 2553

Qy        974 AAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTT 1033

Db       2554 AACCTGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTT 2613

Qy       1034 GTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTA 1093

Db       2614 GTGGATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTG 2673

Qy       1094 ATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGAT 1153

Db       2674 ATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAAC 2733

Qy       1154 CTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATG 1213

Db       2734 TTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATG 2793

Qy       1214 CTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGT 1273

Db       2794 CTGCTGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGAGAGGAGTTTGTGTGC 2853

Qy       1274 GTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGAT 1333

Db       2854 CTCAAATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAG 2913

Qy       1334 GCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGTTT 1390

Db       2914 TCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATC 2973

Qy       1391 TGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTC 1450

Db       2974 CACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTC 3033

Qy       1451 CTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAAC 1510

Db       3034 CTCCTCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGC 3093

Qy       1511 ATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCAC 1570

Db       3094 ATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCAC 3153

Qy       1571 GTGCTTCGCGGGTGCA 1586

Db       3154 CGCCTACATGCGCCCA 3169
```

```
RESULT 20
US-09-864-761-24607
; Sequence 24607, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24607
;  LENGTH: 306
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AL161756.1
;  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   6
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.8
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2
;  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7
;  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   6
;  OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 9.00e-58
;  OTHER INFORMATION: EST HUMAN HIT: BE510841 1  EVALUE 1.00e-46
;  OTHER INFORMATION: NT HIT: AB006590.1, EVALUE 0.00e+00
US-09-864-761-24607

  Query Match             16.1%;  Score 306;  DB 10;  Length 306;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-85;
  Matches  306;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        133 TCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAG 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAATCCATCTTACCCCTGGAGCACGGCTCCATATACATACCTTCCTCCTATGTAGACAG 60

Qy        193 CCACCATGAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCACCATGAAATATCCAGCCATGACATTCTATAGCCCTGCTGTGATGAATTACAGCATTCC 120

Qy        253 CAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTT 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTT 180

Qy        313 GTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTA 240

Qy        373 TGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGT 300

Qy        433 AAACAG 438
              ||||||
Db        301 AAACAG 306


RESULT 21
US-09-864-761-16388
; Sequence 16388, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
```

```
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16388
;   LENGTH: 586
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL   0.46
US-09-864-761-16388

  Query Match             15.9%;  Score 402;  DB 10;  Length 586;
  Best Local Similarity   100.0%;  Pred. No. 5e-84;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         728 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAG 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          33 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAG 92

Qy         788 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGCAC 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          93 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAG 152

Qy         848 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         153 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 212

Qy         908 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         213 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 272

Qy         968 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         273 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 332

Qy        1028 GG 1029
               ||
Db         333 GG 334


RESULT 22
US-09-864-761-32872
; Sequence 32872, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,812
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,406
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,459
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32872
;   LENGTH: 300
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.46
;   OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 6.00e-52
;   OTHER INFORMATION: EST_HUMAN HIT: BE747838.1, EVALUE 6.90e-02
;   OTHER INFORMATION: NT HIT: AB006590.1, EVALUE 0.00e+00
US-09-864-761-32872

  Query Match             15.8%;  Score 300;  DB 10;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-83;
  Matches  300;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         729 GCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGC 788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGC 60

Qy         789 AGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGC 848
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 AGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGC 120

Qy         849 TGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGC 908
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 TGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGC 180

Qy         909 CCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCC 968
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCC 240

Qy         969 TGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCG 1028
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCG 300


RESULT 23
US-10-052-092-2
; Sequence 2, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Bopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02192US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
```

```
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;    LENGTH: 1644
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Synthetic Estrogen Receptor Ligand Binding Domain Fusion
US-10-052-092-2

  Query Match             15.0%;  Score 285.6;  DB 9;  Length 1644;
  Best Local Similarity   63.8%;  Pred. No. 1.3e-78;
  Matches  467;  Conservative    0;  Mismatches  259;  Indels    6;  Gaps    2;

Qy        861 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTCCTGA 920
              ||||||   |||| ||| | ||     || |  |||| |||||||| |||    |
Db        782 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 841

Qy        921 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 977
              |         || |     ||||||| || || || ||||||   |  ||||||||
Db        842 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 901

Qy        978 TGGCCGACAAGGAGTTGGTCACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 1037
              |||| |||| |||| |||| |||||||||| |||||| |||| | | || ||||||||||
Db        902 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 961

Qy       1038 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGACGTCGTTGGATGGAGGTGTTAATGA 1097
              | | | |||    || || || | ||  | ||  |    ||| | ||| |  | ||||
Db        962 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1021

Qy       1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
              | || ||  | |||||||| || || |||||  |  |||||  | ||||||||  |  |
Db        1022 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1081

Qy       1158 TTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
              |  |||||||| |  |||| |||||  |||||| || ||  |||| ||||| |||||||| |
Db        1082 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1141

Qy       1218 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1277
              |||| ||  | ||  |||| ||   | | || || ||      | || | | | ||  |||
Db        1142 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1201

Qy       1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1336
              |  | || ||  |||| |||||  |  ||||| |   |  |  |    |||| | |  ||
Db        1202 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1261

Qy       1337 --GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTGGG 1394
                ||    |     |  |    |  | ||| ||    | || ||  |||| |
Db        1262 TGCAAGAGAAGCACCATATCCACCGAGTGCTGGACAAGATCACAGACACTTTGATCCACC 1321

Qy       1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
              |||| ||||||   ||| |  ||   |||||||   |  ||| ||||  | ||||
Db        1322 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1381

Qy       1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1514
              |  |||| |||||| |||||||   |||||||| |||||||| ||||||   || ||||
Db        1382 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1441

Qy       1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
              ||||||| || |||| ||  | ||||||||||||||||||||||| | ||||||   |
Db        1442 AGTGCAACAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCC 1501

Qy       1575 TTCGCGGGTGCA 1586
              | |  | |  ||
Db        1502 TACATGCGCCCA 1513


RESULT 24
US-10-052-092-19
; Sequence 19, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapy
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
;    LENGTH: 11522
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Vector pER8
US-10-052-092-19

  Query Match             15.0%;  Score 285.6;  DB 9;  Length 11522;
  Best Local Similarity   63.8%;  Pred. No. 4.1e-78;
  Matches  467;  Conservative    0;  Mismatches  259;  Indels    6;  Gaps    2;

Qy        861 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTCCTGA 920
              ||||||   |||| ||| | ||     || |  |||| |||||||| |||    |
Db        880 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 939

Qy        921 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 977
              |         || |     ||||||| || || || ||||||   |  ||||||||
Db        940 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 999

Qy        978 TGGCCGACAAGGAGTTGGTCACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 1037
              |||| |||| |||| |||| |||||||||| |||||| |||| | | || ||||||||||
Db       1000 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1059

Qy       1038 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGACGTCGTTGGATGGAGGTGTTAATGA 1097
              | | | |||    || || || | ||  | ||  |    ||| | ||| |  | ||||
Db       1060 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1119

Qy       1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
              | || ||  | |||||||| || || |||||  |  |||||  | ||||||||  |  |
Db       1120 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1179

Qy       1158 TTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
              |  |||||||| |  |||| |||||  |||||| || ||  |||| ||||| |||||||| |
Db       1180 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1239

Qy       1218 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1277
              |||| ||  | ||  |||| ||   | | || || ||      | || | | | ||  |||
Db       1240 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1299

Qy       1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1336
              |  | || ||  |||| |||||  |  ||||| |   |  |  |    |||| | |  ||
Db       1300 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1359

Qy       1337 --GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTGGG 1394
                ||    |     |  |    |  | ||| ||    | || ||  |||| |
Db       1360 TGCAAGAGAAGCACCATATCCACCGAGTGCTGGACAAGATCACAGACACTTTGATCCACC 1419

Qy       1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
              |||| ||||||   ||| |  ||   |||||||   |  ||| ||||  | ||||
Db       1420 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1479

Qy       1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1514
              |  |||| |||||| |||||||   |||||||| |||||||| ||||||   || ||||
Db       1480 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1539

Qy       1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
```

```
                ||||||| || ||||| ||  | ||||||||||||||||||||||||| | ||||||   |
Db        1540 AGTGCAAGAACGTGGTGTCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCC 1599

Qy        1575 TTCGCGGGTGCA 1586
               | |  | |  ||
Db        1600 TACATGCGCCCA 1611


RESULT 25
US-09-853-033-3
; Sequence 3, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARCETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1983
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: cDS
;   LOCATION: (1)..(1983)
;   OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;   OTHER INFORMATION: Homosapiens Bacteriophage P1
US-09-853-033-3

  Query Match             15.0%;  Score 284;  DB 10;  Length 1983;
  Best Local Similarity   63.7%;  Pred. No. 4.6e-78;
  Matches  466;  Conservative    0;  Mismatches  260;  Indels    6;  Gaps    2;

Qy         861 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 920
               ||||||   |||| || | ||     || |  |||| |||||||| ||     |
Db        1121 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1180

Qy         921 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 977
                 |          || |     ||||||  || || || ||||||   |  ||||||||
Db        1181 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1240

Qy         978 TGGCCGACAAGGAGTTGGTGCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1037
               |||| |||| |||| |||| |||||||||| |||||| ||| | | || ||||||||||
Db        1241 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1300

Qy        1038 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 1097
               |  | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db        1301 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGACATCCTGATGA 1360

Qy        1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
               | || ||  | |||||||| || || ||||| || |||||  | ||||||||  |  |
Db        1361 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGGGAAGCTACTGTTTGCTCCTAACTTGC 1420

Qy        1158 TTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
               |  |||||||| |  |||| ||||| ||||| || ||  |||| ||||| ||||||||  |
Db        1421 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1480

Qy        1218 TGGCAACTACTTCAAGGTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1277
               |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||  |||
Db        1481 TGGCTACATCATCTCGGTTCCGCCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCGTCA 1540

Qy        1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT 1336
               |  | || ||  |||| |||||  |  |||||  |  |  |  |    |||| | |  ||
Db        1541 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1600

Qy        1337  GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1394
                ||   |     |  |   |  | ||| ||    | || ||  |||| |
```

```
Db        1601 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTGATCCACC 1660

Qy        1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
               |||| ||||||   ||| |  ||   |||||||   |  ||| ||||  | |||||
Db        1661 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1720

Qy        1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1514
               |  |||| |||||| |||||||   |||||||||  | ||||| ||||||  || |||||
Db        1721 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAAGAATGGAGCATCTGTACAGCATGA 1780

Qy        1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
               ||||||| || ||||| ||  | |||||||||||||||||||||||| | ||||||   |
Db        1781 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCC 1840

Qy        1575 TTCGCGGGTGCA 1586
               | |  | |  ||
Db        1841 TACATGCGCCCA 1852


RESULT 26
US-09-853-033-7
; Sequence 7, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 1983
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1983)
;   OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;   OTHER INFORMATION: Homosapiens Bacteriophage P1
US-09-853-033-7

  Query Match             14.8%;  Score 280.8;  DB 10;  Length 1983;
  Best Local Similarity   63.4%;  Pred. No. 4.6e-77;
  Matches  464;  Conservative    0;  Mismatches  262;  Indels    6;  Gaps    2;

Qy         861 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 920
               ||||||   |||| || | ||     || |  |||| |||||||| ||     |
Db        1121 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1180

Qy         921 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 977
                 |          || |     ||||||  || || || ||||||   |  ||||||||
Db        1181 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1240

Qy         978 TGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 1037
               |||| |||| |||| |||| |||||||||| |||||| ||| | | || ||||||||||
Db        1241 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1300

Qy        1038 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 1097
               |  | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db        1301 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGACATCCTGATGA 1360

Qy        1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
               | || ||  | |||||||| || || ||||| || |||||  | ||||||||  |  |
Db        1361 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGGGAAGCTACTGTTTGCTCCTAACTTGC 1420

Qy        1158 TTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
               |  |||||||| |  |||| ||||| ||||| || ||  |||| ||||| ||||||||  |
Db        1421 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1480
```

```
Qy       1218 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1277
             |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||  |||
Db       1481 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1540

Qy       1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1336
             |  | || ||  |||| |||||  |  ||||| |  |  |  |    |||| | |  ||
Db       1541 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1600

Qy       1337 --GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1394
                ||    |      |  |    | ||| ||    | || ||  ||||| |
Db       1601 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1660

Qy       1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
             |||| ||||||   ||| |  ||   |||||||   |  ||| ||||  | |||||
Db       1661 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1720

Qy       1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1514
             |  |||| |||||| |||||||   |||||||| |||||||| ||||||  || |||||
Db       1721 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1780

Qy       1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
             ||||||| || ||||| ||  | ||||||||||||||||||  |  | | ||||||  |
Db       1781 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGGCGGCGGACGCCCACCGCC 1840

Qy       1575 TTCGCGGGTGCA 1586
              | |  | |  ||
Db       1841 TACATGCGCCCA 1852


RESULT 27
US-09-853-033-5
; Sequence 5, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;  LENGTH: 1983
;  TYPE: DNA
;  ORGANISM: Artificial sequence
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(1983)
;  OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;  OTHER INFORMATION: Homosapiens-Bacteriophage P1
US-09-853-033-5

  Query Match             14.7%;  Score 279.2;  DB 10;  Length 1983;
  Best Local Similarity   63.3%;  Pred. No. 1.5e-76;
  Matches  463;  Conservative    0;  Mismatches  263;  Indels    6;  Gaps    2;

Qy        861 CCCTGAGCCCCCAGCAGCTACTCCTCACCCTCCTGGAGGCTGAGCCCTACATCCCCGA 42..
              ||||||   |||| ||| | ||     || |  |||| |||||||| |||    |
Db       1121 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGAtGCTGAGCCCCCCATACTCTATT 1180

Qy        921 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATCATGTCCCTGACCAAGT 977
               |         || |     |||||||  || || || ||||||   |  ||||||||
Db       1181 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1240

Qy        978 TGGCCGACAAGGAGTTCGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 1037
              |||| |||| |||| |||| |||||||||| |||||| ||| | | || ||||||||||
Db       1241 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1300

Qy       1038 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 1097
              | | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1301 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1360

Qy       1098 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 1157
              | || ||  | |||||||| || || ||||| |  |||||  | ||||||||  |  |
Db       1361 TTGGTCTGGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1420

Qy       1158 TTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 1217
              |  |||||||| |  |||| ||||| ||||| || ||  |||| ||||| |||||||| |
Db       1421 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1480

Qy       1218 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAACAATATCTCTGTGTCA 1277
              |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||  |||
Db       1481 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1540

Qy       1278 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1336
              |  | || ||  |||| |||||  |  ||||| |  |  |  |    |||| | |  ||
Db       1541 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1600

Qy       1337 --GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1394
                ||    |      |  |    | ||| ||    | || ||  ||||| |
Db       1601 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1660

Qy       1395 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1454
              |||| ||||||   ||| |  ||   |||||||   |  ||| ||||  | |||||
Db       1661 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1720

Qy       1455 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1514
              |  |||| |||||| |||||||   |||||||| |||||||| ||||||  || |||||
Db       1721 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1780

Qy       1515 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1574
              ||||||| || ||||| ||  | ||||||||||||||||||  |  | | ||||||  |
Db       1781 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGGCGGCGGACGCCCACCGCC 1840

Qy       1575 TTCGCGGGTGCA 1586
              | |  | |  ||
Db       1841 TACATGCGCCCA 1852


RESULT 28
US-09-918-995-15134
; Sequence 15134, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15134
;  LENGTH: 451
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(451)
;  OTHER INFORMATION: n = A,T,C or G
US-09-918-995-15134

  Query Match             14.6%;  Score 278;  DB 9;  Length 451;
  Best Local Similarity   90.2%;  Pred. No. 1.5e-76;
  Matches  296;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        731 TCCCGGAGAGAGAGATGTGGGTACCGACTTGTGCTGAAACAGATAAGTGCCGACGAGCAG 790
              |  ||||||||||||||||||||| ||||||| || ||||| ||||||||||| |||
Db         49 TNNAGGAGAGAGAGATGTGGGTACCGACTTGTGCTGAAACAGATAAGTGCCGACGATCAG 108
```

```
Qy        791 CTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTG 850
              ||||||||||||||||||||||| |||||| ||||| ||| ||||||||||||||||||||
Db        109 CTGCACTGTGCCGGCAAGGCCAGGAGAAGGGGCGGTCACACGCCCCGAGTGCGGGAGCTG 168

Qy        851 CTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCC 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        169 CTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGTCC 228

Qy        911 CATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTG 970
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        229 CATGTGCTGATCAGCCGACCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTG 288

Qy        971 ACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGC 1030
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        289 ACCAAGTTGGCCGACAAGGAGTCGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGT 348

Qy       1031 TTTGTGGAGCTCAGCCTGTTCGACCAAG 1058
                 |||         | |    ||||| ||
Db        349 GATGTACCCTCTGGGCACAGCGACCCAG 376


RESULT 29
US-09-737-255-4
; Sequence 4, Application US/09737255
; Patent No. US20020073441A1
; GENERAL INFORMATION:
;  APPLICANT: Rehemtulla, Alnawaz
;  APPLICANT:  Ross, Brian D.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
;  TITLE OF INVENTION:  PROTEOLYTIC ACTIVITY
;  FILE REFERENCE: 11203-003001
;  CURRENT APPLICATION NUMBER: US/09/737,255
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 973
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(973)
US-09-737-255-4

  Query Match             14.1%;  Score 267;  DB 10;  Length 973;
  Best Local Similarity   64.9%;  Pred. No. 6.5e-73;
  Matches  415;  Conservative    0;  Mismatches  225;  Indels    8;  Gaps    1;

Qy        938 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 997
              |||||||  || ||||| ||||||   |   |||||||  | || || | | |||| ||||
Db        184 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACTAGCAGATAGGGAGCTGGTT 243

Qy        998 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAA 1057
              || ||||||| |||||| |||   | || |||||||  |||  | |  ||    || ||
Db        244 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATGATCAG 303

Qy       1058 GTGCGGCTCTTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCA 1117
              || |  ||  | ||| |    ||| |||| |  | |||| || ||  | ||||||||
Db        304 GTCCACTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 363

Qy       1118 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 1177
              || || |||| || |||||| | |||||||  |  |  | |||||||| || | ||
Db        364 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 423

Qy       1178 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTT 1237
              ||||| || |||| ||  |||| ||||||||||||| | || || ||  |     ||||
Db        424 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 483

Qy       1238 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1297
              ||   | | || || ||     |||| | | | ||  |||| |||| ||  |||| |||
Db        484 CGCATGATGAACCTGCAGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 543
```

```
Qy       1298 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1354
              ||| |  |||| |   ||||  ||   |      ||  |   ||    |      | |
Db        544 TCCGGAGTGTACAGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 603

Qy       1355 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1414
                |  | ||| ||    | || ||  |||| |     |||| |||||    ||| |
Db        604 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 663

Qy       1415 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1474
                |    |||||||         |||| ||| | |||||  |  | || |||||  || ||
Db        664 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 723

Qy       1475 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1534
              ||   |||||||| ||||||| ||||   |||||||| ||||| || ||||| ||
Db        724 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTGGTACCC 783

Qy       1535 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1577
                | |||||||||| |||||||| || ||||||||   ||||
Db        784 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 826


RESULT 30
US-09-908-153B-39
; Sequence 39, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Boeril, Roger
;  APPLICANT:  Schopfer, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
;   LENGTH: 2739
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-153B-39

  Query Match             14.0%;  Score 265.4;  DB 9;  Length 2739;
  Best Local Similarity   64.4%;  Pred. No. 4.8e-72;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    8;  Gaps    1;

Qy        938 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 997
              |||||||  || ||||| ||||||   |   |||||||  | || || | | |||| ||||
Db        748 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACTAGCAGATAGGGAGCTGGTT 807

Qy        998 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAA 1057
              || ||||||| |||||| |||   | || |||||||  |||  | |  ||    || ||
Db        808 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATGATCAG 867

Qy       1058 GTGCGGCTCTTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCA 1117
              || |  ||  | ||| |    ||| |||| |  | |||| || ||  | ||||||||
Db        868 GTCCACTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 927

Qy       1118 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 1177
              || || |||| || |||||| | |||||||  |  |  | |||||||| || | ||
Db        928 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 987

Qy       1178 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTT 1237
              ||||| || |||| ||  |||| ||||||||||||| | || || ||  |     ||||
Db        988 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 1047

Qy       1238 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1297
              ||   | | || || ||     |||| | | | ||  |||| |||| ||  |||| |||
Db       1048 CGCATGATGAACCTGCAGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1107
```

```
Db       1048 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1107

Qy       1298 TCCAGTATGTACCC---TCTGGTCACAGCCACCCAGGATGCTGACAGCAGCCGGAAGCTG 1354
              ||| |  ||||| |   ||||  ||   |     ||  |   ||     |       | |

Db       1108 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1167

Qy       1355 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1414
              |   | ||| ||    | || ||  ||||| |     |||| |||||    ||| |

Db       1168 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1227

Qy       1415 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTCTCCCACGTCAGG 1474
              |   ||||||||        ||||| ||| | |||||  |  | || |||||  || ||

Db       1228 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 1287

Qy       1475 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1534
              ||   |||||||||| ||||||||| |||||   |||||||| ||||| || || || ||

Db       1288 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1347

Qy       1535 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1577
              | ||||||||||| |||||||| || ||||||||   ||||
Db       1348 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 1390


RESULT 31
US-09-908-153B-44/c
; Sequence 44, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F
;  APPLICANT:  Beeril, Roger
;  APPLICANT:  Schopfer, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 2739
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-153B-44

  Query Match             14.0%;  Score 265.4;  DB 9;  Length 2739;
  Best Local Similarity   64.4%;  Pred. No. 3.8e-72;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    3;  Gaps    1;

Qy        938 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 997
              |||||||  || ||||| ||||||   |   |||||||  | || || | |||| ||||
Db       1992 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 1933

Qy        998 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGCAGCTCAGCCTGTTCGACCAA 1057
              || ||||||| |||||| ||||   | || ||||||| |||  | |  ||    || ||
Db       1932 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATCATCAG 1873

Qy       1058 GTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTCGCGCTCA 1117
              || |  ||  | ||| |    ||| ||||| |  | ||||| || ||  | ||||||||
Db       1872 GTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 1813

Qy       1118 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGACGGC 1177
              || || ||||| || |||||| | |||||||  |  |  | |||||||| || | ||
Db       1812 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 1753

Qy       1178 AAATGCGTAGAAGGAATTCTCGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTT 1237
              ||||| || ||||| ||  |||| ||||||||||| | || || ||  |     ||||
Db       1752 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 1693

Qy       1238 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1297
              ||   | | || || ||     |||| | | | ||  ||||  |||| ||  |||| |||
Db       1692 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1633

Qy       1298 TCCAGTATGTACCC---TCTGGTCACAGCCACCCAGGATGCTGACAGCAGCCGGAAGCTG 1354
              ||| |  ||||| |   ||||  ||   |     ||  |   ||     |       | |

Db       1632 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1573

Qy       1355 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1414
              |   | ||| ||    | || ||  ||||| |     |||| |||||    ||| |

Db       1572 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1513

Qy       1415 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTCTCCCACGTCAGG 1474
              |   ||||||||        ||||| ||| | |||||  |  | || |||||  || ||

Db       1512 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 1453

Qy       1475 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1534
              ||   |||||||||| ||||||||| |||||   |||||||| ||||| || || || ||

Db       1452 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1393

Qy       1535 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1577
              | ||||||||||| |||||||| || ||||||||   ||||
Db       1392 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 1350


RESULT 32
US-09-908-153B-41
; Sequence 41, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beeril, Roger
;  APPLICANT:  Schopfer, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41
;   LENGTH: 2775
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-153B-41

  Query Match             14.0%;  Score 265.4;  DB 9;  Length 2775;
  Best Local Similarity   64.4%;  Pred. No. 3.8e-72;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    3;  Gaps    1;

Qy        938 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 997
              |||||||  || ||||| ||||||   |   |||||||  | || || | |||| ||||
Db        748 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 807

Qy        998 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGCAGCTCAGCCTGTTCGACCAA 1057
              || ||||||| |||||| ||||   | || ||||||| |||  | |  ||    || ||
Db        808 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATCATCAG 867

Qy       1058 GTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTCGCGCTCA 1117
              || |  ||  | ||| |    ||| ||||| |  | ||||| || ||  | ||||||||
Db        868 GTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 927

Qy       1118 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGACGGC 1177
              || || ||||| || |||||| | |||||||  |  |  | |||||||| || | ||
Db        928 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 987

Qy       1178 AAATGCGTAGAAGGAATTCTCGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGCTTT 1237
              ||||| || ||||| ||  |||| ||||||||||| | || || ||  |     ||||
```

```
Db        988 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 1047

Qy       1238 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1297
              ||    | | || || ||     |||| | | | ||  ||||  |||| ||  |||| |||
Db       1048 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1107

Qy       1298 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1354
              ||| |  |||||  |   ||||  ||    |      ||  |   ||    |       | |
Db       1108 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1167

Qy       1355 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGACGGCATC 1414
                |   | ||| ||    | || ||  ||||| |     |||| |||||    ||| |
Db       1168 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1227

Qy       1415 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1474
                |    ||||||||        ||||| ||| | ||||  |  | || |||||  || ||
Db       1228 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 1287

Qy       1475 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1534
              ||   |||||||||| |||||||| |||||   |||||||| ||||| || || || ||
Db       1288 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1347

Qy       1535 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1577
                | ||||||||||| |||||||| || ||||||||   ||||
Db       1348 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 1390


RESULT 43
US-09-908-154B-43/c
; Sequence 43, Application US/09908154B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beerli, Roger
;  APPLICANT:  Schopfer, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,154B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
;   LENGTH: 2775
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-154B-43

  Query Match             14.0%;  Score 265.4;  DB 9;  Length 2775;
  Best Local Similarity   64.4%;  Pred. No. 4.8e-72;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    3;  Gaps    1;

Qy        948 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 997
              |||||||  || |||||  ||||||   |   |||||||  | || || | |||| ||||
Db       2028 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 1969

Qy        998 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAA 1057
              || ||||||| |||||| ||||   | || ||||||| |||  | |  ||    || ||
Db       1968 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATGATCAG 1909

Qy       1058 CTCCGGCTCTTTGGAGACCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCA 1117
              ||  |  ||  | ||| |    ||| ||||| |  | ||||| || ||  | ||||||||
Db       1908 GTCCACTTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 1849

Qy       1118 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 1177
              || || ||||| || |||||| | ||||||||  |  |  | |||||||| || | ||
Db       1848 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 1789
```

```
Qy       1178 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGTTT 1237
              ||||| || |||| ||  |||| ||||||||||| | || || ||  |     ||||
Db       1788 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 1729

Qy       1238 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1297
              ||    | | || || ||     |||| | | | ||  ||||  |||| ||  |||| |||
Db       1728 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1669

Qy       1298 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1354
              ||| |  |||||  |   ||||  ||    |      ||  |   ||    |       | |
Db       1668 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1609

Qy       1355 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGACGGCATC 1414
                |   | ||| ||    | || ||  ||||| |     |||| |||||    ||| |
Db       1608 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1549

Qy       1415 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1474
                |    ||||||||        ||||| ||| | ||||  |  | || |||||  || ||
Db       1548 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 1489

Qy       1475 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1534
              ||   |||||||||| |||||||| |||||   |||||||| ||||| || || || ||
Db       1488 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1429

Qy       1535 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1577
                | ||||||||||| |||||||| || ||||||||   ||||
Db       1428 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 1386


RESULT 44
US-09-864-761-6718
; Sequence 6718, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
```

```
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6718
;   LENGTH: 599
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
US-09-864-761-6718

  Query Match             12.2%;  Score 232.4;  DB 10;  Length 599;
  Best Local Similarity   99.6%;  Pred. No. 3.5e-62;
  Matches  233;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1481 AGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTAT 1540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 AGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTAT 425

Qy       1541 GACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAAGTCCTCCATCACG 1600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAAGTCCTCCATCACG 485

Qy       1601 GGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAG 1660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 GGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAG 545

Qy       1661 TCTCAGTGACGCCTGGCCCTGAGGTGAACTGGCCCACAGAGGTCACAAGCTGAA 1714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTCAGTGACGCCTGGCCCTGAGGTGAACTGGCCCACAGAGGTCACAGGCTGAA 599


RESULT 35
US-09-864-761-23452
; Sequence 23452, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,412
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23452
;   LENGTH: 183
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
;   OTHER INFORMATION: NT HIT: X99101.1, EVALUE 6.00e-99
;   OTHER INFORMATION: EST_HUMAN HIT: AI174706.1, EVALUE 2.50e-00
;   OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 4.00e-30
US-09-864-761-23452

  Query Match              9.6%;  Score 183;  DB 10;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-47;
  Matches  183;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1483 TAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGA 1542
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGA 60

Qy       1543 CCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAAGTCCTCCATCACGGG 1602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGGTGCAAGTCCTCCATCACGGG 120

Qy       1603 GTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTC 1662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTC 180

Qy       1663 TCA 1665
             |||
Db        181 TCA 183


RESULT 36
US-09-962-436-268
; Sequence 268, Application: US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
;  APPLICANT: Soppet, Daniel
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-75
;  CURRENT APPLICATION NUMBER: US/09/962,436
;  CURRENT FILING DATE:  2001-09-25
```

```
;  PRIOR APPLICATION NUMBER: US/60/235,082
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/234,924
;  PRIOR FILING DATE: 2000-09-25
;  NUMBER OF SEQ ID NOS: 568
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 268
;   LENGTH: 2218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-962,436-268

  Query Match              7.0%;  Score 133.8;  DB 10;  Length 2218;
  Best Local Similarity   49.9%;  Pred. No. 6.5e-41;
  Matches  392;  Conservative    0;  Mismatches  487;  Indels    6;  Gaps    2;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | ||||| ||||| | ||    || || || || |||||||| ||    || |||
Db        406 CGCCTCTGCCTGGTCTGTGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 465

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || | || || ||||| || || || | ||| || ||    | || || | |||||
Db        466 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCG 525

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 694
              ||  | ||  ||||| ||| ||    | |||||| ||||||||||||||| |
Db        526 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 585

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTAC 754
              |||||       ||||| ||| | || ||    ||     ||    || |     ||
Db        586 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGT 645

Qy        755 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 814
              | |    | |    || | | | | | | |       |       ||
Db        646 GGGCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTC 705

Qy        815 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAG 874
              |||   ||  ||   |   |   |     | | |   | | |  |||   |   |
Db        706 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCCCCAGTGAATGCACTG 765

Qy        875 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 934
              |       ||     | | ||| |||||  || | |||    ||     || | |  |
Db        766 GTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGC 825

Qy        935 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 994
              |       |   | || ||   |   ||||    |  |   |||   ||| |
Db        826 CCTGATGGGCACCTCCCAGCCGTGGCTA---CCCTCTGTGACCTCTTTGACCGAGAGATT 882

Qy        995 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 1054
              ||   || ||||||||||||||||  || || |||||      |||    ||||  |||
Db        883 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 942

Qy        1055 CAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGC 1114
               ||  ||    | || |||||  ||||||||||| |  || |||||      |||
Db        943 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGC 1002

Qy        1115 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAG 1174
               ||| |    |  | |  |||   ||| |||   ||  | || |||||     || |||
Db        1003 TCACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTAGTCCTGGA---TGAAGAG 1059

Qy        1175 GGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGG 1234
               ||| | | | ||  || |  |||| |    |     || ||| |              ||
Db        1060 GGGGCACGGGCAGCTGGCCTGGGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGCGG 1119

Qy        1235 TTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGCCATGATCCTGCTC 1294
               | | |   |  || | |    | || ||| |     | |||||| ||  |||   |
Db        1120 CTGCAGGCCCTGCCGCTGGAGCGAGAGGAGTATGTTCTACTAAAGGCCTTGGCCCTTGCC 1179

Qy        1295 AATTC 1299
               |||||
Db        1180 AATTC 1184

RESULT 47
US-10-211-239-7
; Sequence 7, Application US/10211239
; Publication No. US20030103965A1
; GENERAL INFORMATION:
;  APPLICANT: Junq, Birqit
;  APPLICANT:  Krant, No. US20030103965Albert
;  APPLICANT:  Mueller, Stefan
;  TITLE OF INVENTION: Method for Identifying Substances Which Positively
;  TITLE OF INVENTION:  Influence Inflammatory Conditions
;  FILE REFERENCE: 0652.2340001
;  CURRENT APPLICATION NUMBER: US/10/211,239
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/315,775
;  PRIOR FILING DATE: 2001-08-30
;  PRIOR APPLICATION NUMBER: EP 01119003.0
;  PRIOR FILING DATE: 2001-08-06
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 2402
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-211-239-7

  Query Match              6.8%;  Score 129;  DB 9;  Length 2402;
  Best Local Similarity   48.7%;  Pred. No. 2.2e-29;
  Matches  382;  Conservative    0;  Mismatches  400;  Indels    3;  Gaps    1;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | ||||| ||||| | ||    || || || || |||||||| ||    || |||
Db        583 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 642

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || | || || ||||| || || || | ||| || ||    | || || | |||||
Db        643 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCG 702

Qy        635 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 694
              ||  | ||  ||||| ||| ||    | |||||| ||||||||||||||| |
Db        703 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 762

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTAC 754
              |||||       ||||| ||| | || ||    ||     ||    || |     ||
Db        763 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGT 822

Qy        755 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 814
              | |    | |    || | | | | | | |       |       ||
Db        823 GGGCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTC 882

Qy        815 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAG 874
              |||   ||  ||   |   |   |     | | |   | | |  |||   |   |
Db        883 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCA 942

Qy        875 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 934
              | | |   |     ||  ||| |     |   |     ||     ||||
Db        943 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCA 1002

Qy        935 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 994
              |  ||         |   | |     ||   ||||     | |   |||   ||| |
Db        1003 GGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATT 1062

Qy        995 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 1054
              ||   || ||||||||||||||||  || || |||||      |||    ||||  |||
Db        1063 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 1122

Qy        1055 CAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGC 1114
               ||  ||    | || |||||  ||||||||||| |  || |||||      |||
Db        1123 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGC 1182

Qy        1115 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAG 1174
               ||| |    |  | |  |||   ||| |||   ||  | || |||||     || |||
Db        1183 TCACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTAGTCCTGGA---TGAAGAG 1239
```

```
Qy       1175 GGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGG 1234
             ||| | | | ||  || |  |||| |     |      || ||| |           ||
Db       1240 GGGGCACGGGCAGCTGGCCTGGGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGCGG 1299

Qy       1235 TTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTC 1294
              | |  |   |    || | |   | || ||| |     | |||||| ||  |||   |
Db       1300 CTGCAGGCCCTGCGGCTGGAGCGAGAGGAGTATGTCTACTAAAGGCCTTGGCCCTTGCC 1359

Qy       1295 AATTC 1299
              |||||
Db       1360 AATTC 1364


RESULT 38
US-10-037-270-454
; Sequence 454, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje I.
;  TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784C1P2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 454
;   LENGTH: 2539
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (97)..(1755)
US-10-037-270-454

  Query Match             6.8%;  Score 129;  DB 9;  Length 2539;
  Best Local Similarity   48.7%;  Pred. No. 2.2e-29;
  Matches  382;  Conservative    0;  Mismatches  400;  Indels    3;  Gaps    1;

Qy        515 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 574
              | | |||||   |||||  | ||    || || || || |||||||| ||    || |||
Db        712 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 771

Qy        575 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 634
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db        772 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGACCATCGACTACAGCTGTCCC 831

Qy        635 GCTACAAATCACTCTACAATCGATAAAAACCCCCCCAAGAGCTGCCAGGCCTCCCACTT 694
              ||  | ||  |||||  |||  ||     | ||||||  ||||||||||||||||  |
Db        832 GCCTCCAACGAGTGTGAGATCACCAAGCGCAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 891

Qy        695 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGAGAGAGACATCTGCCTAC 754
                |||||       ||||| ||| | |||   ||    ||    ||  |     ||

Db        892 ACCAAGTGCCTGCGGCTGGGCATGCTCAAGCAGGCAGTGCGCCTGGACGAGGTGTAGGGGT 951

Qy        755 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 814
              | |    |  |    || |  |  | || | | | |  |       |         ||
Db        952 GGGCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTC 1011

Qy        815 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGGACGCCCTGAGCCCCGAG 874
              |||   ||  ||  |  |  |  | | |  |  |||       |          |
Db       1012 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCA 1071

Qy        875 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 934
              | | |    |     ||  |||  |  |    |    |          ||     ||||
Db       1072 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCGCA 1131

Qy        935 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 994
              |  ||         |   | |      ||    ||||     |  |    |||   ||| |
Db       1132 GGCCCTGATGGGCACCTCCCAGCCGTGGCTAGCCTCTGTGACCTCTTTGACCGAGAGATT 1191

Qy        995 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 1054
              ||   ||  ||||||||||||||||  || || |||||      |||    ||||  |||
Db       1192 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 1251

Qy       1055 CAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGC 1114
              ||  ||    |  || |||||  ||||||||||||| |  || ||||  ||      ||||
Db       1252 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGC 1311

Qy       1115 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAC 1174
              ||| |    |  |  |  ||||   ||| |||   ||  | || ||||    || |||
Db       1312 TCACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTAGTCCTGGA---TGAAGAG 1368

Qy       1175 GGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGG 1234
              ||| | | | ||  || |  |||| |     |      || ||| |           ||
Db       1369 GGGGCACGGGCAGCTGGCCTGGGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGCGG 1428

Qy       1235 TTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTC 1294
              | |  |   |    || | |   | || ||| |     | |||||| ||  |||   |
Db       1429 CTGCAGGCCCTGCGGCTGGAGCGAGAGGAGTATGTCTACTAAAGGCCTTGGCCCTTGCC 1488

Qy       1295 AATTC 1299
              |||||
Db       1489 AATTC 1493


RESULT 39
US-10-008-739A-1
; Sequence 1, Application US/10008739A
; Patent No. US20020161194A1
; GENERAL INFORMATION:
;  APPLICANT: Pfizer Inc.
;  APPLICANT:  Castleberry, Tessa A.
;  APPLICANT:  Lu, Bihong
;  APPLICANT:  Owen, Thomas A.
;  APPLICANT:  Smock, Steven L.
;  TITLE OF INVENTION: The Canine Androgen Receptor
;  FILE REFERENCE: PC10893AGPR
;  CURRENT APPLICATION NUMBER: US/10/008,739A
;  CURRENT FILING DATE:  2002-04-15
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 3577
;   TYPE: DNA
;   ORGANISM: Canine
US-10-008-739A-1

  Query Match             5.1%;  Score 97.2;  DB 9;  Length 3577;
  Best Local Similarity   51.7%;  Pred. No. 2.6e-19;
  Matches  332;  Conservative    0;  Mismatches  288;  Indels   22;  Gaps    4;

Qy        528 TCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGG 587
              |||| | ||| | || || | | ||||||||||      | |||| | | || || |
Db       1646 TCTGCGGTCATCAACGCTTCTGCCTGTCACTATGGAGTCTTCACTTGTGGAAGCTGCAAAG 1705
```

```
Qy        588 CCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGT 647
              ||| |||||||||  |  | ||||  |  |  | ||  | ||| |     | |||| | |
Db       1706 TCTTCTTTAAAAGAGCCGCTGAAGGAAACAGAAGTACCTGTGTGCCAGCAGAAATGATT 1765

Qy        648 GTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACG 707
              |||| ||||||||||  |||  | || |  || |    | || || || ||||| || || |
Db       1766 GTACCATCGATAAATTCCGAAGGAAAAATTGTCCATCTTGTCGCCTCCGGAAATGCTATG 1825

Qy        708 AAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGC--G 765
              |||  || |||      | | | |     |  | ||| |   ||| | |       ||| |
Db       1826 AAGCAGGGATGACTCTGGGAGCCCCGGAAGCTAAAGAAACTGGGGAATCTGAAACTGCAAG 1885

Qy        766 GAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGG 825
              ||  |||| |   | |     || | || |       ||     |    |  || |||
Db       1886 AGGAAGGAGAGGCTTCCAATGTCACCAGCCCCACTGAGGAGCCAACCCAGAAGCTGACGG 1945

Qy        826 CCACCGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCT 885
              | | |   ||  ||       ||              | |||||          | ||
Db       1946 TGTCACACATTGAAGGCTATGAGTG--------    ----TCAGCCCATCTTTCTGAATGTC 1992

Qy        886 CACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCAC 945
              |      |   |    || ||     |||||||  ||    |  ||    ||||  | |
Db       1993 CTTCAAGCCATCGAGCCAGGCGTGGTGTGTCTGGACA-TGACAACAACCAGCCCGACTTC 2051

Qy        946 CGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGAT 1005
              |   ||  || || |      ||| |   | |||  || | | || | |||||  || |
Db       2052 CTTTGCAGCCTTGCTCTCTAGCCTTAATGAATTGGGTGAAAGGCAGCTTGTACATGTGGT 2111

Qy       1006 CAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCT 1065
              ||  |||||||||    | || ||||     | ||   | ||   ||||  ||    |
Db       2112 CAAGTGGGCCAAGGCCTTGCCGGGCTTCCGCAACCTGCACGTGGATGACCAGATGGCAGT 2171

Qy       1066 CTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCG------CTCAAT 1119
              | |  ||  ||  ||||||| | |   |  || | |   ||  |||||       | | |
Db       2172 CATTCAGTACTCCTGGATGGGGCTCATGGTGTTTGCCATGGGCTGGCGATCCTTCACCAA 2231

Qy       1120 TGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCT 1161
              || | ||  || | | |||||  ||| || || || || || ||| |
Db       2232 TGTCAACTCCAGGATGCTCTACTTCGCCCCTGACCTGGTTTT 2273


RESULT 40
US-09-997-267-1
; Sequence 1, Application US/09997267
; Patent No. US20020165381A1
; GENERAL INFORMATION:
;  APPLICANT: AHRENS FATH, ISABELLE
;  APPLICANT:  HAENDLER, BERNARD
;  TITLE OF INVENTION: HUMAN ANDROGEN RECEPTOR VARIANTS
;  FILE REFERENCE: SCH-1794
;  CURRENT APPLICATION NUMBER: US/09/997,267
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: 60/255,078
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1329
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-997-267-1

  Query Match             5.0%;  Score 95.2;  DB 9;  Length 1329;
  Best Local Similarity   51.2%;  Pred. No. 6.2e-19;
  Matches  432;  Conservative    0;  Mismatches  298;  Indels   18;  Gaps    4;

Qy        520 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGGAAGG 579
              ||||   ||||  | ||| | || || || | ||||||||||||      | |||| | |
Db        243 CTGCCTGATCTGTGGAGATGAAGCTTCTGGGTGTCACTATGGAGCTCTCACATGTGGAAG 302

Qy        580 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 639
              || |||| ||| || ||||||  |  | ||||  |  |  | ||  |  |     |
Db        303 CTGCAAGTCTTCTTCAAAAGAGCCGCTGAAGGAAACAGAAGTACCTGTGCGCCAGCAG 362

Qy        640 AAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 699
              |||| | || || || |||||  ||| | || |  || |    | || || ||||||||
Db        363 AAATGATTGCACTATTGATAAATTCCGAAGGAAAAATTGTCCATCTTGTCGTCTTCGGAA 422

Qy        700 GTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCT 759
              ||||| |||| || |||     | | | |     |   ||| |  ||| | |
Db        423 ATGTTATGAAGCAGGGATGACTCTGGGAGCCCCGGAAGCTGAAGAAACTTGGTAATCTGAA 482

Qy        760 TGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAG 819
              | | | ||   || || | |  |||| |  ||      | |    | |||   |
Db        483 ACTACAG-GAGGAAGGAGAGGCTTCCAGCACCACCAGCCCCACTGAGGAGACAACCCAGA 541

Qy        820 TGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCT 879
              | | |   | | |       ||        ||          | |||||
Db        542 AGCTGACAGTGTCACACATTGAAGGCTATGAATG----------TCAGCCCATCTTTCTG 591

Qy        880 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCC 939
              | ||  |      |   |   ||  |      ||||||| || | |  ||    |||
Db        592 AATGTCCTGGAAGCCATTGAGCCAGGTGTAGTGTGTGCTGGACA-CGACAACAACCAGCC 650

Qy        940 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACA 999
              |  | ||   || || || |      ||| |   |  ||| || |   || | |||||
Db        651 CGACTCCTTTGCAGCCTTGCTCTCTAGCCTCAATGAACTGGGAGAGAGACAGCTTGTACA 710

Qy       1000 CATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGT 1059
              | || |||  |||||||||     | || |||||      |  |   | ||  |||||  |
Db        711 CGTGGTCAAGTGGGCCAAGGCCTTGCCTGGCTTCCGCAACTTACACGTGGACGACCAGAT 770

Qy       1060 GCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCG-----  1113
              |    || |  ||  ||  ||||||| | |   |  || | |   ||  |||||
Db        771 GGCTGTCATTCAGTACTCCTGGATGGGGCTCATGGTGTTTGCCATGGGCTGGCGATCCTT 830

Qy       1114 CTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCT 1161
              | | | || | || || | | ||||  ||| || || |||| ||| |
Db        831 CACCAATGTCAACTCCAGGATGCTCTACTTCGCCCCTGACCTGGTTTT 878


Search completed: June 16, 2003, 13:06:45
Job time : 291.725 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

SM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 07:16:17 ; Search time 224.63 Seconds
(without alignments)
9686.913 Million cell updates/sec

Title: US-08-826-361A-1
Perfect score: 1434
Sequence: 1 ATGAATTACAGCATTCCCAG.....AGAACCCACAGTCTCAGTGA 1434

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1434	100.0	1740	9	US-10-198-785-1	Sequence 1, Appli
2	1434	100.0	1740	10	US-09-815-068A-1	Sequence 1, Appli
3	395.8	27.6	2764	10	US-09-893-666A-1	Sequence 1, Appli
4	390.8	27.3	1788	10	US-09-953-933-1	Sequence 1, Appli
5	390.8	27.3	4963	10	US-09-281-674-9	Sequence 9, Appli
6	390.8	27.3	4963	10	US-09-892-227-9	Sequence 9, Appli
7	389.2	27.1	2092	9	US-10-052-092-6	Sequence 6, Appli
8	389.2	27.1	2092	9	US-09-952-680A-6	Sequence 6, Appli
9	389.2	27.1	6450	9	US-10-096-710-2	Sequence 2, Appli
10	389.2	27.1	6450	9	US-10-081-563-1	Sequence 1, Appli
11	389.2	27.1	6450	9	US-10-052-092-1	Sequence 1, Appli
12	389.2	27.1	6450	9	US-10-052-092-7	Sequence 7, Appli
13	389.2	27.1	6610	9	US-10-052-092-20	Sequence 20, Appl
14	388	27.1	1788	9	US-10-052-092-24	Sequence 24, Appl
15	385.2	26.9	6608	9	US-10-198-846-11305	Sequence 11305, A
16	366.2	25.5	1978	9	US-10-052-092-23	Sequence 23, Appl
17	326.6	22.8	2100	9	US-10-052-092-5	Sequence 5, Appli
18	306.8	21.4	3300	9	US-10-259-864-3	Sequence 3, Appli
19	302	21.1	586	10	US-09-864-761-16388	Sequence 16388, A
20	300	20.9	300	10	US-09-864-761-32872	Sequence 32872, A
21	295.6	20.6	1644	9	US-10-052-092-2	Sequence 2, Appli
22	285.6	19.9	11522	9	US-10-052-092-19	Sequence 19, Appl
23	284	19.8	1983	10	US-09-853-033-3	Sequence 3, Appli
24	280.8	19.6	1983	10	US-09-853-033-7	Sequence 7, Appli
25	279.2	19.5	1983	10	US-09-853-033-5	Sequence 5, Appli
26	278	19.4	451	9	US-09-918-995-15184	Sequence 15184, A
27	277	19.3	973	10	US-09-737-255-4	Sequence 4, Appli
28	265.4	18.5	2739	9	US-09-908-153B-39	Sequence 39, Appl
c 29	265.4	18.5	2739	9	US-09-908-153B-44	Sequence 44, Appl
30	265.4	18.5	2775	9	US-09-908-153B-41	Sequence 41, Appl
c 31	265.4	18.5	2775	9	US-09-908-153B-43	Sequence 43, Appl
32	203	14.2	306	10	US-09-864-761-24607	Sequence 24607, A
33	203	14.2	597	10	US-09-864-761-7903	Sequence 7903, Ap
34	189	13.2	599	10	US-09-864-761-6718	Sequence 6718, Ap
35	183	12.8	183	10	US-09-864-761-23452	Sequence 23452, A
36	134.8	9.4	2218	10	US-09-962-436-268	Sequence 268, App
37	129	9.0	2402	9	US-10-211-239-7	Sequence 7, Appli
38	129	9.0	2539	9	US-10-037-270-454	Sequence 454, App
39	97.2	6.8	3577	9	US-10-008-739A-1	Sequence 1, Appli
40	95.2	6.6	1329	9	US-09-997-267-1	Sequence 1, Appli
41	95.2	6.6	3715	10	US-09-880-107-2300	Sequence 2300, Ap
42	95.2	6.6	4393	9	US-10-198-846-10819	Sequence 10819, A
43	95	6.6	2582	9	US-10-202-846-1	Sequence 1, Appli
44	93.6	6.5	1171	9	US-09-997-267-3	Sequence 3, Appli
45	89	6.2	6244	10	US-09-281-674-8	Sequence 8, Appli
46	89	6.2	6244	10	US-09-892-227-8	Sequence 8, Appli
47	88.6	6.2	2481	9	US-10-211-239-8	Sequence 8, Appli
48	88.6	6.2	2481	10	US-09-853-386-89	Sequence 89, Appl
49	88.6	6.2	2626	12	US-10-002-600-45	Sequence 45, Appl
50	88.2	6.2	2029	10	US-09-954-456-1892	Sequence 1892, Ap
51	88.2	6.2	2573	12	US-10-002-600-44	Sequence 44, Appl
52	88	6.1	465237	10	US-09-933-267A-1	Sequence 1, Appli
53	86.4	6.0	939	9	US-10-153-668-219	Sequence 219, App
54	86.4	6.0	939	9	US-10-153-668-221	Sequence 221, App
55	86.4	6.0	2681	9	US-10-153-668-223	Sequence 223, App
56	86.4	6.0	4789	9	US-10-153-668-225	Sequence 225, App
57	84.8	5.9	1263	9	US-10-153-668-217	Sequence 217, App
58	84.8	5.9	1884	10	US-09-853-386-62	Sequence 62, Appl
59	84.8	5.9	1884	12	US-10-005-169-1	Sequence 1, Appli
60	84.8	5.9	3273	9	US-10-259-864-5	Sequence 5, Appli
61	84.8	5.9	6127	10	US-09-916-145-1	Sequence 1, Appli
62	84.8	5.9	6522	15	US-09-917-806A-1546	Sequence 1546, Ap
63	84.8	5.9	7252	9	US-10-001-486B-1	Sequence 1, Appli
64	83.2	5.8	1606	10	US-09-853-386-55	Sequence 55, Appl
65	83.2	5.8	3802	10	US-09-853-386-46	Sequence 46, Appl
66	83.2	5.8	3802	10	US-09-853-386-78	Sequence 78, Appl
67	83.2	5.8	5115	10	US-09-853-386-57	Sequence 57, Appl
68	81	5.6	2799	10	US-09-935-368-2	Sequence 2, Appli
69	80.8	5.6	1577	10	US-09-797-727-1	Sequence 1, Appli
70	80.8	5.6	2469	10	US-09-853-386-90	Sequence 90, Appl
71	79.2	5.5	3421	10	US-09-853-386-14	Sequence 14, Appl
72	79.2	5.5	3427	9	US-10-171-581-253	Sequence 253, App
73	79.2	5.5	3427	10	US-09-853-386-4	Sequence 4, Appli
74	78.4	5.5	4400	10	US-09-853-386-47	Sequence 47, Appl
75	78.4	5.5	4400	12	US-10-005-169-3	Sequence 3, Appli
76	75.2	5.2	489	10	US-09-960-352-7483	Sequence 7483, Ap
77	74.4	5.2	2154	10	US-09-853-386-24	Sequence 24, Appl
78	74.4	5.2	2247	10	US-09-853-386-29	Sequence 29, Appl
79	74.4	5.2	2353	10	US-09-853-386-23	Sequence 23, Appl
80	74.4	5.2	2458	10	US-09-853-386-3	Sequence 3, Appli
81	74.4	5.2	2967	10	US-09-954-456-521	Sequence 521, App
82	73.4	5.1	1597	9	US-10-153-668-215	Sequence 215, App
83	73	5.1	785	10	US-09-853-386-60	Sequence 60, Appl
84	71.2	5.0	1375	9	US-10-239-804-7	Sequence 7, Appli
85	71.2	5.0	1399	9	US-10-239-804-8	Sequence 8, Appli
86	71.2	5.0	9127	9	US-10-239-804-9	Sequence 9, Appli
87	71.2	5.0	9151	9	US-10-239-804-10	Sequence 10, Appl
88	70.2	4.9	355	9	US-10-081-563-3	Sequence 3, Appli
89	70.2	4.9	42366	9	US-10-081-563-25	Sequence 25, Appl
90	69.6	4.9	2666	10	US-09-853-386-2	Sequence 2, Appli
91	66.6	4.6	1594	10	US-09-964-824A-559	Sequence 559, App
92	65.4	4.6	1431	9	US-10-188-721-2	Sequence 2, Appli

93	65.4	4.6	2218	10	US-09-880-107-3411	Sequence 3411, Ap
94	65.4	4.6	2220	9	US-10-182-232-6	Sequence 6, Appli
95	64.8	4.5	1404	10	US-09-965-703-62	Sequence 62, Appl
96	64.8	4.5	2711	10	US-09-952-559-2	Sequence 2, Appli
97	64.4	4.5	481	10	US-09-864-761-16345	Sequence 16345, A
98	64	4.5	440	9	US-09-918-995-32194	Sequence 32194, A
99	62.2	4.3	549	9	US-09-991-936-1417	Sequence 1417, Ap
100	62	4.3	1759	10	US-09-962-446-300	Sequence 300, App
101	62	4.3	5642	12	US-10-002-600-53	Sequence 53, Appl
102	60.8	4.2	198	10	US-09-965-703-41	Sequence 41, Appl
103	60.8	4.2	1247	10	US-09-965-703-70	Sequence 70, Appl
104	60.8	4.2	1288	10	US-09-965-703-1	Sequence 1, Appli
105	60.8	4.2	1290	10	US-09-965-697-4	Sequence 4, Appli
106	60.8	4.2	1542	10	US-09-965-703-58	Sequence 58, Appl
107	60.8	4.2	2005	12	US-10-142-373-1	Sequence 1, Appli
108	60.4	4.2	987	10	US-09-965-703-21	Sequence 21, Appl
109	60.4	4.2	1398	10	US-09-909-672-1	Sequence 1, Appli
110	60.2	4.2	2511	10	US-09-919-497-36	Sequence 36, Appl
111	60	4.2	392	9	US-10-052-092-4	Sequence 4, Appli
112	60	4.2	1247	9	US-10-052-092-3	Sequence 3, Appli
113	58.8	4.1	463	9	US-09-918-995-15693	Sequence 15693, A
114	58.8	4.1	1740	9	US-10-161-804-26	Sequence 26, Appl
115	57	4.0	2751	10	US-09-728-422-2	Sequence 2, Appli
116	56	3.9	486	9	US-09-918-995-13740	Sequence 13740, A
117	56	3.9	1123	10	US-09-965-703-26	Sequence 26, Appl
118	55.8	3.9	2066	10	US-09-895-840-1	Sequence 1, Appli
119	55.8	3.9	2126	9	US-10-295-370-1	Sequence 1, Appli
120	55.2	3.8	571	9	US-10-066-543-458	Sequence 458, App

ALIGNMENTS

```
RESULT 1
US-10-198-785-1
; Sequence 1, Application US/10198785
; Publication No. US20030022224A1
; GENERAL INFORMATION:
;  APPLICANT: OLYMPUS OPTICAL CO., LTD.
;  TITLE OF INVENTION: Method of detecting binding reaction between protein and test sub
;  FILE REFERENCE: ZE1-02S0538
;  CURRENT APPLICATION NUMBER: US/10/198,785
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: JP/2001-220444
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: JP/2001-221963
;  PRIOR FILING DATE: 2001-07-23
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1740
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (99)..(1688)
US-10-198-785-1

  Query Match             100.0%;  Score 1434;  DB 9;  Length 1740;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 317

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGCCACCTTTCTCCTTTAGTGGTCCATGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGCCACCTTTCTCCTTTAGTGGTCCATGGC 377

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 437

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 497

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 557

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 617

Qy        361 AAAAGAAGCATTCAAGCACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 677

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 737

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 797

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 857

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 917

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 977

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1037

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1097

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1157

Qy        901 CTCATCTTTGCTTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CTCATCTTTGCTTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 1217

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1277

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1337

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1397

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1457

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1517
```

```
Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1577

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1637

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1691


RESULT 2
US-09-815-068A-1
; Sequence 1, Application US/09815068A
; Patent No. US20020072076A1
; GENERAL INFORMATION:
;  APPLICANT: OLYMPUS OPTICAL CO., LTD.
;  TITLE OF INVENTION: Method for testing a substance interacting with hormone receptor
;  FILE REFERENCE: 00S0496P
;  CURRENT APPLICATION NUMBER: US/09/815,068A
;  CURRENT FILING DATE:  2001-03-22
;  PRIOR APPLICATION NUMBER: JP/11-209860
;  PRIOR FILING DATE: 1999-07-23
;  PRIOR APPLICATION NUMBER: JP/2000-163475
;  PRIOR FILING DATE: 2000-05-31
;  PRIOR APPLICATION NUMBER: JP/2000-163476
;  PRIOR FILING DATE: 2000-05-31
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;  LENGTH: 1740
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (99)..(1688)
US-09-815-068A-1

  Query Match           100.0%;  Score 1434;  DB 10;  Length 1740;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1434;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 317

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 377

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 437

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 497

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 557

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTC 617

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 677

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTCTTACCAACTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTCTTACCAACTGGGA 737

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGACAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 ATGGTGAAGTGTGGCTCCCGGAGAGACAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 797

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 857

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 917

Qy        661 CAGGCTGAGCCGCCCCATGTGTTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 GAGGCTGAGCCGCCCCATGTGTTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 977

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1037

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1097

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1157

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1217

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1277

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1337

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1397

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1457

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATG 1517

Qy       1261 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 GAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAG 1577

Qy       1321 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 ATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGGGTCCGAGTGCAGC 1637

Qy       1381 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 CCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTCTCAGTGA 1691


RESULT 3
US-09-893-666A-1
; Sequence 1, Application US/09893666A
; Patent No. US20020129386A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: YAMASHITA, ICHIRO
;  TITLE OF INVENTION: High estrogen-sensitive medaka fish
;  FILE REFERENCE: 21021?US 620 7249-0
;  CURRENT APPLICATION NUMBER: 09/893,666A
;  CURRENT FILING DATE:  2002-02 12
;  PRIOR APPLICATION NUMBER: JP 2000 247729
;  PRIOR FILING DATE: 2000 08 17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;  LENGTH: 2764
;  TYPE: DNA
;  ORGANISM: Oryzias latipes
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (211)..(1935)
;  OTHER INFORMATION:
US 09 893 666A 1

  Query Match             27.6%;  Score 395.8;  DB 10;  Length 2764;
  Best Local Similarity   62.9%;  Pred. No. 8e-112;
  Matches  704;  Conservative    0;  Mismatches  467;  Indels   48;  Gaps    4;

Qy        274 GATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGG 333
              ||  | |  |||||||| || ||||||| ||||| || || || |||||||| || |||
Db        619 GACACGCGTTTCTGCGCCGTGTGCAGCGACTACGCCTCTGGGTACCACTATGGGGTCTGG 678

Qy        334 TCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATT 393
              || ||||| || || |||||||| || || || ||||| || || || ||||| |||||
Db        679 TCTTGTGAGGGCTGCAAGGCCTTCTTCAAGAGGAGCATCCAGGGTCACAATGACTATATG 738

Qy        394 TGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGC 453
              || ||||| || |||||||| || || || | ||| ||| | ||||||||||||||| ||
Db        739 TGCCCAGCGACCAATCAGTGCACTATTGACAGAAATCGGAGGAAGAGCTGCCAGGCTTGT 798

Qy        454 CGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG---------------      494
              || ||| ||||||||||||||||||||||||| ||||  | ||
Db        799 CGTCTTAGGAAGTGTTACGAAGTGGGAATGATGAAAGGCGGTGTGCGCAAGGACCGCATT 858

Qy        495 ---CTCCCGGAGAGAGACATGTGGGTACCGCCTTTGTGCGGACAGACAGAAGTGCCGAC 549
                  | ||| | || | | | ||    || |||      || | |    ||   |
Db        859 CGCATTTTACGGCGTGACAAACGGCGGACAGGCGTTCGGTGATGGAGACAAGGTTGTAAAG 918

Qy        550 GAGCAGCTGCACTGTGCCG -----          GCAAGGCCAAGAGAAGTGGCGGCCAC 594
              | ||| |||     | |                | ||   ||  || | |  || ||
Db        919 GGTCAGGAGCATAAAACGGTGCATTATGATGGAAGGAAACGCAGCAGCACAGGAGCAGGA 978

Qy        595 GCGCCCCGAGTGCGGGAGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACC 654
              |     |||   | |      |||   |  ||  |  ||     |||  ||||  |
Db        979 GGAGGAGCAGGAGGAGGAAGACTGTCTGTGACCAGCATACCTCCTGAGCAGGTGCTGCTC 1038

Qy        655 CTCCTGGAGGCTGAGCCGCCCCCATGTGCTGATCAGCGCC      CCAGTGCGCCCTTC 708
              |||||  |||  |    |||||   | ||   |  || |        ||    || | |
Db        1039 CTCCTTCAGGGCCCGAGCCCCCGATACTCTGCTCGTGTCAGAAGTTGAGCCGACCGTAC 1098

Qy        709 ACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATG 768
              |||||| | |||||||||     || ||||   |||| |||||||| |||| ||||||
Db        1099 ACCGAGGTCACCATGATGACCCTGCTCACCAGCATGGCAGACAAGGAGCTGGTCCACATG 1158

Qy        769 ATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGG 828
              |||  |||||||||||||| | || || ||| || ||||   ||||  ||| || |||| |
Db        1159 ATCGCCTGGGCCAAGAAGCTCCCAGGTTTCTGCAGCTGTCCCTGCACGATCAGGTGCTG 1218

Qy        829 CTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGCGGCTGATGTGGCGCTCAATTGAC 888
              ||  |||||||||  ||| |||||||| | ||||| || || || ||| | || ||  ||
Db        1219 CTGCTGGAGAGCTCGTGGCTGGAGGTGCTCATGATCGGCCTCATTTGGAGGTCCATCCAC 1278

Qy        889 CACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGC 948
                |||||  |||||||||||||| | ||| ||   | |||||||| |||||||  | |||
Db        1279 TGTCCCGGGAAGCTCATCTTTGCACAAGACCTCATCCTGGACAGGAATGAGGGAGACTGC 1338

Qy        949 GTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGTTTCGAGAG 1008
              || |||||  ||   ||| |||||  |||||||| |||||  ||| ||||   | || || | |
Db        1339 GTGGAAGGCATGACGGAGATCTTCGACATGCTGCTGGCCACTGCTTCCCGCTTCCGTGTG 1398

Qy        1009 TTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGT 1068
               | ||||||  |||    | ||||    ||||  |||| || || ||  | |||||  ||| ||
Db        1399 CTCAAACTCAAACCTGAGGAATTCGTCTGCCTCAAAGCTATTATTTTACTCAACTCCGGT 1458

Qy        1069 ATGTACGCTCTGGTCACAGCGACCCAGGATGC   TGACAGCAGCCGGAAGCTGGCTCAC 1125
                 |   || |   ||| |  ||| ||| |    | ||| ||||  |  | |       |
Db        1459 GCTTTTCTTTCTGCACCGGCACCATGGAGCCACTTCACAACAGCGCGGCGGTTCAGAGC 1518

Qy        1126 TTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCC 1185
               ||||| || || | || || ||  |  ||      ||      ||  ||  ||  ||
Db        1519 ATGCTGGACACCATCACAGACGCACTCATTCATTACATCAGTCAGTCGGGTTACTTGGCC 1578

Qy        1186 CAGCAGCAATCCATCGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCG 1245
               ||| ||||  | |  || | |||  | || ||  |||| || |||||| |||||||   |
Db        1579 CAGGAGCAGGCGAGACGGCAGGCCCAGCTGCTCCTGCTGCTCTCCCACATCAGGCACATG 1638

Qy        1246 AGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTAT 1305
               || ||||| |||||||| || ||    || ||||||||||| ||     |||||  | |||
Db        1639 AGCAACAAAGGCATGGAGCACCTCTACAGCATGAAGTGCAAGAACAAAGTCCCTCTTTAT 1698

Qy        1306 GACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGC 1344
               ||||| || ||||||||||||  ||||||||   || | |
Db        1699 GACCTCCTACTGGAGATGCTCGATGCCCACCGCCTGCAC 1737


RESULT 4
US-09-853-033-1
; Sequence 1, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;  LENGTH: 1788
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(1788)
US-09-853-033-1

  Query Match             27.4%;  Score 390.8;  DB 10;  Length 1788;
  Best Local Similarity   62.4%;  Pred. No. 2.2e-110;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   89;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| ||  || || ||||  || || || || || || ||||||||||
Db        538 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATATGGAGTC 597

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||| || |||
Db        598 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 657

Qy        391 ATTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || |||||| || || || | ||||||||| || | ||||||||||||||
Db        658 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCT 717

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||  ||||    || | |||  ||
Db        718 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 777

Qy        511 TGTGG-----------GTACCGCCTTGTGCGGACAGACAGAAGTGCCGACGAGCAGCTG 558
```

```
Db        778 AGAGGAGGGAGAATGTTGAAACACAAGGCCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 837

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 608
              | | |     | || |||| || |  ||    ||| |  ||
Db        838 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 897

Qy        609 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                     | |    |||       ||||||   |||| ||| | | ||     || |
Db        898 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 957

Qy        658 CTGGAGGCTGAGCCCGCCCATGTGCTGATC---AGCCGCCCCAGTGCCCCCTTCACCCAC 714
              |||| |||||||| |||    |      |         || |     |||||||  ||
Db        958 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCGTTCAGTGAA 1017

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||   |  ||||||||  |||| |||| |||| |||| |||||||||| |
Db       1018 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1077

Qy        775 TGGGCCAAGAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCTGCTCTTG 834
              ||||| |||| | | || ||||||||||||  | | |||    || || || |  ||  |
Db       1078 TGGGCGAAGAGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1137

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1138 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1197

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTACAA 954
              || |||||  | |||||||  |  |  |  |||||||| |  |||| |||| |||||
Db       1198 GGCAAGCTACTGTTTGCTCCTAACTTCGTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1257

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              ||  ||  |||| ||||| |||||||| |||| ||  |  | ||  |||| ||   | | ||
Db       1258 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGCTTCCGCATGATGAAT 1317

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              ||  ||     | || | | | ||  |||  | || ||  |||| |||||  |  |||||
Db       1318 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1377

Qy       1075 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
              |  |  |  |  |||| | |  ||    |     |  |    |  | |||
Db       1378 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1437

Qy       1132 AACGGCCTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1438 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1497

Qy       1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
              ||   |  ||| |||||  | ||||  |  |||| ||||| ||||||   ||||||
Db       1498 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1557

Qy       1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
              || |||||||| ||||||  || ||||||||||| || |||| ||  | ||||||||
Db       1558 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1617

Qy       1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1351
              ||||||||||||||| | ||||||   || |  | |  ||
Db       1618 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 1657


RESULT 5
US-09-281-674-9
; Sequence 9, Application US/09281674
; Patent No. US20020077307A1
;   GENERAL INFORMATION:
;        APPLICANT: Gossen, Manfred
;                   Bujard, Hermann
;                   Salfeld, Jochen
;                   Voss, Jeffrey
;        TITLE OF INVENTION: Methods for Regulating Gene Expression
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Lahive & Cockfield
;             STREET: 60 State Street, Suite 510
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: ASCII text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/281,674
;             FILING DATE: 30-Mar-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/479,306
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/260,452
;             FILING DATE: 14-JUN-1994
;             APPLICATION NUMBER: 08/076,327
;             FILING DATE: 14-JUN-1993
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Giulio A. DeConti, Jr.
;             REGISTRATION NUMBER: 31,503
;             REFERENCE/DOCKET NUMBER: BBI-013CP3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 227-7400
;             TELEFAX: (617) 227-5941
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4963 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        ORIGINAL SOURCE:
;             ORGANISM: Human cytomegalovirus
;        IMMEDIATE SOURCE:
;             CLONE: pUHD BGR4
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-281-674-9

  Query Match             27.4%;  Score 390.8;  DB 10;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 3.9e-110;
  Matches  699;  Conservative    0;  Mismatches  382;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db       1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1122

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || |||||||| || | ||||||||||||||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGCAGAGAGAA 510
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACGGA 1242

Qy        511 TGTGC----------GTACCGCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
              | ||          | | | |    | |||| |    | |  | | |
Db       1243 AGAGGAGGGAGAATGTTGAAACACAAGGCCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 608
                  | | |     | || |||| || |  ||    ||| |  ||
Db       1303 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        609 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                     | |    |||       ||||||   |||| ||| | | ||  | || |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCATGGCCTTG 1422
```

```
Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||| |||||||| |||    |       |          || |      ||||||| ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||    |  |||||||||  |||| |||| |||| |||| ||||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCTTCCATGATCAGGTCCACCTTCTA 1602

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1603 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCCTCTGGCGCTCCATGGAGCACCCA 1662

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 954
              |  |||||  | |||||||  |  |  |  |||||||| |  |||| |||||  |||||
Db       1663 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1722

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| |||||  |||||||| |||||  ||  | ||  |||| ||   | | ||
Db       1723 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1782

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              || ||     | || | | | ||  ||||  | || ||  |||| |||||  |  |||||
Db       1783 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1842

Qy       1075 CCTCTGGTCACAGCGACCCAGCATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
              |  | | |    |||| | |  ||    ||     |      |  |    | |||
Db       1843 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1902

Qy       1132 AACCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
              ||    | || ||  |||||  |     |||| ||||||   ||| |  ||   ||||||
Db       1903 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1962

Qy       1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
              ||   |  ||| |||||  | |||||  |  |||| ||||||  ||||||   |||||||
Db       1963 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 2022

Qy       1252 AAGGGCATGCAACATCTGCTCAACATGAAGTGCAAAAATGTGGTTCCCAGTGTATGACCTG 1311
              || |||||||| ||||||  || ||||||||||| || ||||| ||  | ||||||||||
Db       2023 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 2082

Qy       1312 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 1351
              |||||||||||||||| | ||||||   || |  | |  ||
Db       2083 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2122


RESULT 6
US-09-892-227-9
; Sequence 9, Application US/09892227
; Patent No. US20020152487A1
;   GENERAL INFORMATION:
;        APPLICANT: Gossen, Manfred
;                   Bujard, Hermann
;                   Salfeld, Jochen
;                   Voss, Jeffrey
;        TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
;                            Transcriptional
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Lahive & Cockfield
;             STREET: 60 State Street, Suite 510
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109 1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC DOS/MS DOS
;             SOFTWARE: ASCII text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/892,227
;             FILING DATE: 25-Jun-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/163,269
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/076,327
;             FILING DATE: 14 JUN 1993
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Giulio A. DeConti, Jr.
;             REGISTRATION NUMBER: 31,503
;             REFERENCE/DOCKET NUMBER: BBI-013CP2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 227-7400
;             TELEFAX: (617) 227-5941
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4963 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        ORIGINAL SOURCE:
;             ORGANISM: Human cytomegalovirus
;        IMMEDIATE SOURCE:
;             CLONE: pUHD BGR4
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-892-227-9

  Query Match             27.3%;  Score 390.8;  DB 10;  Length 4963;
  Best Local Similarity   62.4%;  Pred. No. 3.9e-110;
  Matches  699;  Conservative    0;  Mismatches  482;  Indels   89;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db       1003 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 1062

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| |||||  || || |||||||| || || ||||| ||||||||||||||| || |||
Db       1063 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAACGACATAACGACTAT 1122

Qy        391 ATTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||
Db       1123 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1182

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||| || | |||  ||
Db       1183 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1242

Qy        511 TGTGG--------  GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
                | ||          | | | |      | |||| |     | |  | |  |
Db       1243 AGAGGAGGGAGAATGTTCAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 1302

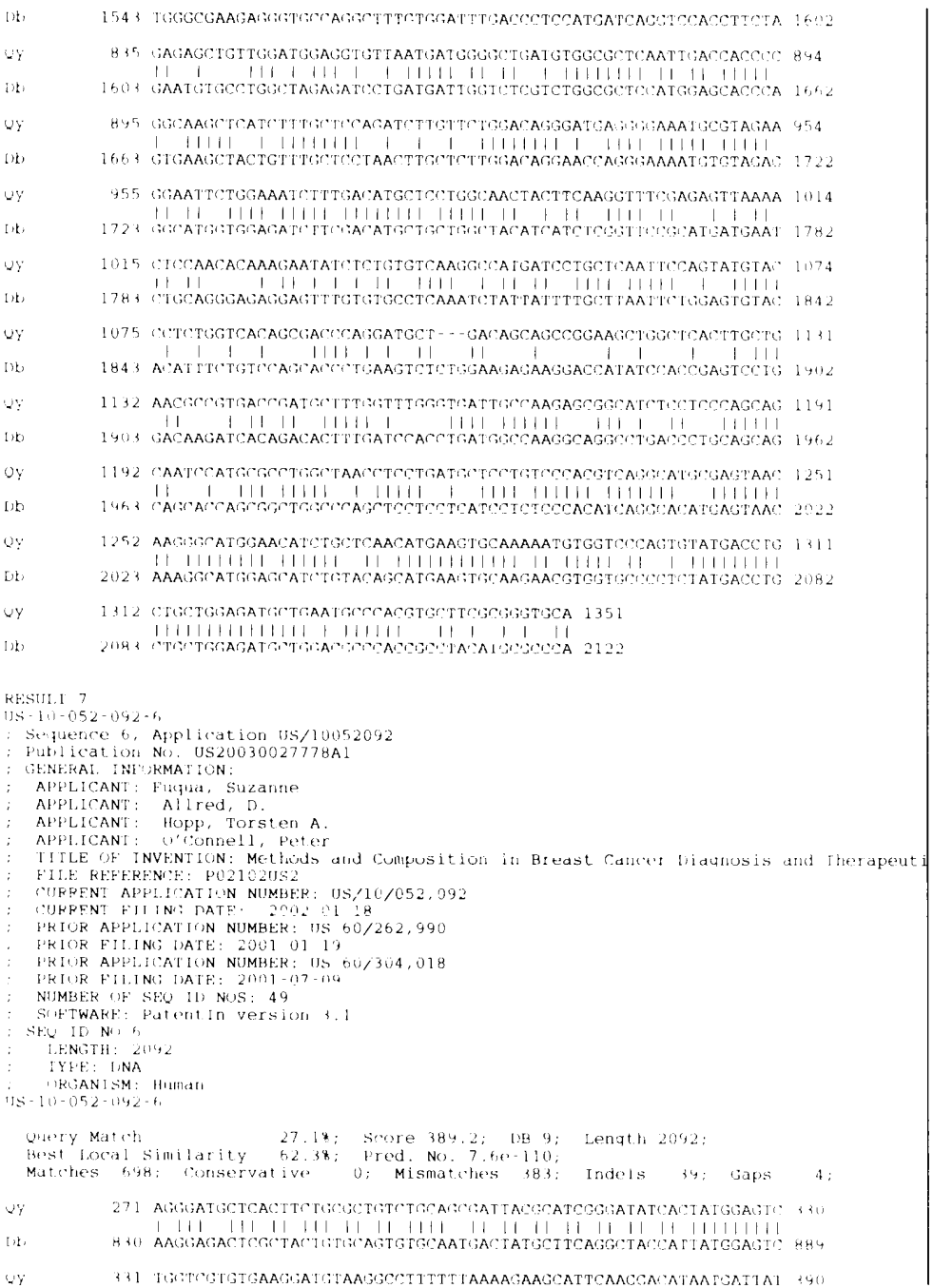
Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG --  -     608
                |  | |     | || |||  || |  ||   |||  |  ||
Db       1303 GTGGGGTCTGCCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1362

Qy        609 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                      | |    |||       ||||||   |||| ||| | ||  |  || | |
Db       1363 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCATGGCCTTG 1422

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||| |||||||| |||    |       |          || |      ||||||| ||
Db       1423 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1482

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||    |  |||||||||  |||| |||| |||| |||| ||||||||||| |
Db       1483 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1542

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1543 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCTTCCATGATCAGGTCCACCTTCTA 1602
```



RESULT 7
US-10-052-092-6
; Sequence 6, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
; APPLICANT: Fuqua, Suzanne
; APPLICANT: Allred, D.
; APPLICANT: Hopp, Torsten A.
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
; FILE REFERENCE: P02102US2
; CURRENT APPLICATION NUMBER: US/10/052,092
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/262,990
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/304,018
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 2092
; TYPE: DNA
; ORGANISM: Human
US-10-052-092-6

Query Match 27.1%; Score 389.2; DB 9; Length 2092;
Best Local Similarity 62.3%; Pred. No. 7.6e-110;
Matches 698; Conservative 0; Mismatches 383; Indels 39; Gaps 4;



RESULT 8

```
US 09 952 680A 9
; Sequence 9, Application US/09952680A
; Publication No. US20040087239A1
; GENERAL INFORMATION:
;  APPLICANT: Stanton, Marty
;  APPLICANT:  Epstein, David
;  APPLICANT:  Hamaguchi, No. US20040087239Aluko
;  TITLE OF INVENTION: Target Activated Biosensor and Methods of Using Same
;  FILE REFERENCE: 23239 501
;  CURRENT APPLICATION NUMBER: US/09/952,680A
;  CURRENT FILING DATE:  2001 09 13
;  PRIOR APPLICATION NUMBER: 60/232,454
;  PRIOR FILING DATE: 2000 09 13
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 2092
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US 09 952 680A-9

  Query Match             27.1%;  Score 489.2;  DB 9;  Length 2092;
  Best Local Similarity   62.3%;  Pred. No. 7.6e-110;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || ||||||||| 
Db        830 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 889

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        890 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 949

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || |||||||| || | ||||||||||||||| 
Db        950 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1009

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| || 
Db       1010 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGTGGGATACGAAAAGACCGA 1069

Qy        511 TGTGG---  ---GTACCGCCTTGTGCGGAGACAGAGAAGTCCCGACGAGCAGCTG 558
              | |||          | | | |     | |||| |     | | | | |
Db       1070 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGCTGAA 1129

Qy        559 CACTGTGCCCGCAAGCCAAGACAAGTGGCCGCCACGCGCCCCGAGTGCG---          608
                | | |     | || |||  || |  ||    ||| |   ||
Db       1130 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCGCTCATGATCAAA 1189

Qy        609   ---       GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                | |    |||       ||||||   |||| ||| | ||     || |
Db       1190 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1249

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC--AGCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||| |||||||| |||    |  |      || |      |||||| ||
Db       1250 TTGGATGTGACCCCCCATACTCTATTCCCAGTATCATCCTACCAGACCCTTCAGTGAA 1309

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||   |  ||||||||  |||| |||| |||| |||| ||||||||| |
Db       1310 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1369

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              ||||| |||| | | || |||||||||||  | | |||    || || || | |  || |
Db       1370 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1429

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1430 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1489

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAA 954
              |  |||||  | |||||||| |  | |  | |||||||| |  |||| ||||| ||||| 
Db       1490 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1549

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| ||||| |||||||| ||||| ||  | ||  |||| ||   | || |
Db       1550 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1609

Qy       1015 CTCCAACACAAAGAATATCTCTGTCTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              || ||    | || | | | ||  |||| | || ||  |||| ||||| |  |||||
Db       1610 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTGCTTAATTCTGGAGTGTAC 1669

Qy       1075 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
              | | | |   |||| | |  ||  ||   |  | |   |  | ||| 
Db       1670 ACATTCTGTCCAGCACCCTGAAGTCTCTGGAAGACAAGGACCATATCCACCGAGTCCTG 1729

Qy       1132 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
              ||    | || ||  |||| |     |||| ||||||   ||| | |  ||   ||||||
Db       1730 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1789

Qy       1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
              ||   | ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1790 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1849

Qy       1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
              || ||||||| |||||  || |||||||||||| || ||||| ||  | |||||||||
Db       1850 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1909

Qy       1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1351
              |||||||||||||| | ||||||   || |  | |  ||
Db       1910 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 1949


RESULT 9
US-10-096-710-2
; Sequence 2, Application US/10096710
; Patent No. US20020164581A1
; GENERAL INFORMATION:
;  APPLICANT: Chang, Chawnshang
;  TITLE OF INVENTION: METHODS FOR IDENTIFYING COMPOUNDS THAT
;  TITLE OF INVENTION:  DO NOT ATTENUATE THE PROTECTIVE EFFECTS OF ESTROGEN
;  FILE REFERENCE: 21108.0010U2
;  CURRENT APPLICATION NUMBER: US/10/096,710
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/275,229
;  PRIOR FILING DATE: 2001-03-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 6450
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:/note
;   OTHER INFORMATION: synthetic construct
US 10-096-710 2

  Query Match             27.1%;  Score 489.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 1.4e-109;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || ||||||||| 
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || |||||||| || | ||||||||||||||| 
Db       1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| || 
Db       1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGTGGGATACGAAAAGACCGA 1137
```

```
Qy        511 TGTGG------------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
              | ||              | | | |       | |||| |     | |  | | |
Db       1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGCAGGGCAGGGGTGAA 1197

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 608
              |  | |     | || |||| || |  ||    ||| |   ||
Db       1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1257

Qy        609 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                       | |   |||       ||||||  |||| ||| | ||     || |
Db       1258 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1317

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||| |||||||| |||    |         |           || |     |||||||  ||
Db       1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||   |  |||||||| |||| |||| |||| |||| ||||||||| |
Db       1378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCGCGCTCTTG 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  || |
Db       1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | |||| || ||  | |||||| || || |||||
Db       1498 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1557

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 954
              |  |||||  | |||||||| |  | |  |  |||||||| |  |||| |||| ||||
Db       1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| |||||  |||||||| |||||  ||  | ||  |||| ||   | | ||
Db       1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCCCATCATGAAT 1677

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              || ||     | || | | | ||  |||  | || ||  |||| |||||  |  |||||
Db       1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTCGAGTGTAC 1737

Qy       1075 CCTCTGGTCACAGCGACCCACGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
              |  |  | |    |||| | |  ||   ||    |      |  |     |  | |||
Db       1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy       1132 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
              ||   | || ||  |||| |    |||| |||||  ||| |  ||   |||||
Db       1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy       1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||    |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy       1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
              || |||||||| ||||||  || ||||||||||||| || ||||| ||  | |||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy       1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1351
              |||||||||||||| | ||||||   || |  | |  ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 10
US-10-081-563-1
; Sequence 1, Application US/10081563
; Publication No. US20020187495A1
; GENERAL INFORMATION:
;  APPLICANT: Wake Forest University
;  APPLICANT:  Herrington, David M.
;  APPLICANT:  Howard, Timothy D.
;  APPLICANT:  Hawkins, Gregory A.
;  APPLICANT:  Meyers, Deborah A.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS OF ESTROGEN RECEPTOR ALPHA ASSOCIATED WITH
;  TITLE OF INVENTION:  FAVORABLE HDL CHOLESTEROL RESPONSE TO HORMONE REPLACEMENT THERAP
;  FILE REFERENCE: 9151-15
;  CURRENT APPLICATION NUMBER: US/10/081,563
;  CURRENT FILING DATE:  2002-02-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 6450
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (361)..(2148)
;   OTHER INFORMATION:
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: X03635
;   DATABASE ENTRY DATE: 1993-09-12
;   RELEVANT RESIDUES: (1)..(6450)
US-10-081-563-1

  Query Match             27.1%;  Score 389.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 1.4e-109;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGCAGTC 957

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| |||| || || |||||||| || || |||| ||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACGGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || |||||||| || | ||||||||||||||||
Db       1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGACAGACA 510
              ||||| || || || || ||||||||||||||| ||||  ||||    || | |||  ||
Db       1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1137

Qy        511 TGTGG------------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
              | ||              | | | |       | |||| |     | |  | | |
Db       1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGCAGGGCAGGGGTGAA 1197

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 608
              |  | |     | || |||| || |  ||    ||| |   ||
Db       1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1257

Qy        609 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                       | |   |||       ||||||  |||| ||| | ||     || |
Db       1258 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1317

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||| |||||||| |||    |         |           || |     |||||||  ||
Db       1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||   |  |||||||| |||| |||| |||| |||| ||||||||| |
Db       1378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCGCGCTCTTG 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  || |
Db       1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | |||| || ||  | |||||| || || |||||
Db       1498 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1557

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 954
              |  |||||  | |||||||| |  | |  |  |||||||| |  |||| |||| ||||
Db       1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy        955 GGAATTCTGGAAATCTTTCACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| |||||  |||||||| |||||  ||  | ||  |||| ||   | | ||
```

```
Db       1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
             || ||     | || | | | ||  |||| | || ||  |||| |||||  | |||||
Db       1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1737

Qy       1075 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
              |  |  | |    |||| | |  ||   ||     |      | |    |  | |||
Db       1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy       1132 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
              ||    | || ||  ||||| |     |||| ||||||   ||| |  ||   ||||||
Db       1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy       1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
              ||    |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy       1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
              || ||||||| ||||||  || ||||||||||| || |||||  ||  | |||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCCTCTATGACCTG 1977

Qy       1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCCGCGGGTGCA 1351
              |||||||||||||||| | ||||||   || |  | |  ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 11
US-10-052-092-1
; Sequence 1, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/404,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 6450
;   TYPE: DNA
;   ORGANISM: Human
US-10-052-092-1

  Query Match             27.1%;  Score 389.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 1.4e-109;
  Matches  698;  Conservative    0;  Mismatches  483;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || ||||||| || || ||||| || || |||||||| || | ||||||||||||||||
Db        1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| ||||    || | | |||  ||
Db        1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1137

Qy        511 TGTGG--        ---GTACCGCCTTGTGCGGAGA?AGAGAAGTGCCGACGAGCAGCTG 558
```

```
                | ||           | | | |      | |||| |     | |  | |  |
Db       1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGCAGGGCAGGGGTGAA 1197

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 608
                | | |      | || |||| || |  ||    |||  |   ||
Db       1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1257

Qy        609 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                      | |    |||       ||||||   |||| ||| | ||      || |
Db       1258 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1317

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||| |||||||| |||     |       |          || |     |||||||  ||
Db       1318 TTGGATGCTGAGCCCCCCATACTCTATTCCCAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||    |  |||||||  |||| |||| |||| |||| ||||||||| |
Db       1378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  || |
Db       1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |   ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1498 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1557

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 954
              |  ||||  | ||||||||  |  |  |  |||||||| |  |||| |||||  |||||
Db       1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| |||||  |||||||| |||||  ||  | ||  |||| ||   | | ||
Db       1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677

Qy        1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
               || ||     | || | | | ||  |||| | || ||  |||| |||||  | |||||
Db       1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1737

Qy        1075 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
               |  |  | |    |||| | |  ||   ||     |      | |    |  | |||
Db       1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy        1132 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
               ||    | || ||  ||||| |     |||| ||||||   ||| |  ||   ||||||
Db       1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy        1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
               ||    |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy        1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
               || ||||||| ||||||  || ||||||||||| || |||||  ||  | |||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCCTCTATGACCTG 1977

Qy        1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCCGCGGGTGCA 1351
               |||||||||||||||| | ||||||   || |  | |  ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017


RESULT 12
US-10-052-092-7
; Sequence 7, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
```

```
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 6450
;   TYPE: DNA
;   ORGANISM: Human
US-10-052-092-7
```

```
  Query Match             27.1%;  Score 389.2;  DB 9;  Length 6450;
  Best Local Similarity   62.3%;  Pred. No. 1.4e-109;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 1017

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db       1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 1077

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db       1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 1137

Qy        511 TGTGG----------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
              | ||          | | | |      | |||| |     | |  | |  |
Db       1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGCAGGGCAGGGTGAA 1197

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 608
                | | |     | || |||| || |  ||    ||| |   ||
Db       1198 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 1257

Qy        609 ----------GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACGCTC 657
                       | |    |||       ||||||   |||| ||| | ||     || |
Db       1258 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 1317

Qy        658 CTGGAGGCTGAGGCCGCCTATGTGGTCATC---AGCGCCCCCAGTGCGCCCTTCACCGAG 714
               |||| |||||||| ||    |      |      || |    |||||||  ||
Db       1318 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1377

Qy        715 GCCTCCATGATGATGTGCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||    | ||||||||  |||| |||| |||| |||| ||||||||| |
Db       1378 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 1437

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGTTGAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1438 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1497

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | |||| || ||  | |||||||| || || |||||
Db       1498 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 1557

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 954
              |  |||||  | |||||||| |  |  |  |||||||| |  |||| |||| |||||
Db       1558 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 1617

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| |||| |||||||| |||| ||  | ||  |||| ||   | | ||
Db       1618 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 1677

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              || ||     | || | | | ||  |||  | || ||  |||| |||| |  ||||
Db       1678 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 1737

Qy       1075 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
                | | | |  |||| | | ||   ||    |     | |   |  | |||
Db       1738 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 1797

Qy       1132 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCAGCAG 1191
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1798 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1857

Qy       1192 CAATCCATGGCGCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       1858 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 1917

Qy       1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
              || |||||||| ||||||  || |||||||||||| || |||| ||  | |||||||||
Db       1918 AAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 1977

Qy       1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1351
              |||||||||||||| | ||||||   || |  | |  ||
Db       1978 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2017
```

```
RESULT 13
US-10-052-092-20
; Sequence 20, Application Us/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapy
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
;   LENGTH: 6610
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Vector pCI-n GLI-HEGO
US-10-052-092-20
```

```
  Query Match             27.1%;  Score 389.2;  DB 9;  Length 6610;
  Best Local Similarity   62.3%;  Pred. No. 1.4e-109;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || |||||||||
Db       2422 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 2481

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db       2482 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 2541

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || |||| || || |||||||| || | ||||||||||||||||
Db       2542 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 2601

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||| || | ||| ||
Db       2602 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 2661

Qy        511 TGTGG----------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
              | ||          | | | |      | |||| |     | |  | |  |
Db       2662 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGCAGGGCAGGGTGAA 2721

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCG---------- 608
                | | |     | || |||| || |  ||    ||| |   ||
```

```
Db       2722 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 2781

Qy        609            GGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                         | |    |||       ||||||   |||| ||| | ||     || |
Db       2782 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACCGCCGACCAGATGGTCAGTGCCTTG 2841

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC--AGCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||| |||||||| ||    |       |          || |     |||||||  ||
Db       2842 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 2901

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||   |  ||||||||  |||| |||| |||| |||| ||||||||||| |
Db       2902 GCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGATCAAC 2961

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  || |
Db       2962 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 3021

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       3022 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCA 3081

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 954
              |  |||||  | |||||||| |  | |  |  |||||||| |  |||| ||||| |||||
Db       3082 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAG 3141

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| ||||| |||||||| ||||| ||  | ||  |||| ||   | | ||
Db       3142 GGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGATGAAT 3201

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              || ||    | || | | | ||  |||  | || ||  |||| ||||  |  |||||
Db       3202 CTGCAGGGAGAGCAGTTTGTGTGCCTCAAATCTATTATTTTGCTTAATTCTGGAGTGTAC 3261

Qy       1075 CCTCTGGTCACAGCGACCCAGGATGCT--GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
                |  |  |  |    |||| | |  ||  ||    |  |  |  |  | |||
Db       3262 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGAGTCCTG 3321

Qy       1132 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
               ||    | || ||  |||| |      |||| ||||||   ||| |  ||   ||||||
Db       3322 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 3381

Qy       1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   |||||||
Db       3382 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGTAAC 3441

Qy       1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
              || ||||||| |||||||  || ||||||||||||| || |||||  ||  | |||||||||
Db       3442 AAACGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTG 3501

Qy       1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1351
              |||||||||||||| | ||||||   || |  | |  ||
Db       3502 CTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCA 3541


RESULT 14
US-10-052-092-24
; Sequence 24, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torston A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;  LENGTH: 1788
;  TYPE: DNA
;  ORGANISM: Pig
US-10-052-092-24

  Query Match             27.1%;  Score 488;  DB 9;  Length 1788;
  Best Local Similarity   62.4%;  Pred. No. 1.6e-109;
  Matches  695;  Conservative    0;  Mismatches  380;  Indels   89;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || ||||||||
Db        548 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCCTCAGGCTACCATTATGGAGTT 547

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| || || || |||||||||| || || |||| |||||||||||||||||| ||
Db        598 TGGTCTTGCGAGGGCTGTAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAATGACTAC 657

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||||| || ||||| ||||| ||||| ||| || | |||||||| ||||||
Db        658 ATGTGTCCAGCTACCAACCAGTGCACAATTGATAAGAACAGGAGGAAGAGCTGTCAGGCC 717

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCCGA      502
              ||||| || || |||| ||||||||||| ||| |||| | ||    ||||
Db        718 TGCCGGCTACGCAAGTGCTACGAAGTGGGCATGATGAAAGGGGGGATACGGAAAGACCGG 777

Qy        503 ---GAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
                 ||| ||||     | | | |       | |||| |     | |  | | | |
Db        778 AGAGGAGGGAGAATGTTGAAGCACAAGCGCCAGAGAGATGATGGAGAGGGCAGGAATGAA 837

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGCCCCCG---      ---- 602
                   ||      | || |||| || | ||   |||  |
Db        838 GCGGTGCCCCCTGGAGACATGAGATCTGCCAACCTTTGGCCAAGCCCTCTCTTGATTAAA 897

Qy        603 -----AGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                  | | |    | | |      ||||||   |||| ||| | ||    || |
Db        898 CACACTAAGAAGAACAGCCCGGTCTTGTCCCTGACAGCCGACCAGATGATCAGTGCCTTG 957

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTTCACCGAG 714
              |||||||||||||| |||    |       |          || |     ||| |||  |||
Db        958 TTGGAGGCTGAGCCCCCCATAATCTATTCCGAGTATGATCCTACCAGACCCCTCAGTGAG 1017

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || ||||||   |  ||||||||  | || |||| |||| ||||||||||||||| |
Db       1018 GCTTCAATGATGGGCTTGCTGACCAACCTGGCAGACAGGGAGCTGGTACACATGATCAAC 1077

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              ||||| |||| | | || || ||| |||  | |||||     || ||||||  ||  ||
Db       1078 TGGGCAAAGAGGGTGCCAGGATTTTTGGATTTAAGCCTCCATGATCAAGTGCATCTTCTG 1137

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1138 GAATGTGCCTGGCTAGAGATCCTCATGATTGGTCTTGTCTGGCGCTCCATGGAGCACCCA 1197

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAA 954
              || |||||  | |||||||  |  |  | |||||||| |  |||| || || || ||
Db       1198 GGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAACCAGGGCAAGTGTGTCGAG 1257

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              |||||  |||| ||||||||||| | ||||| ||  | ||  | || ||   | | ||
Db       1258 GGAATGGTGGAGATCTTTGACATGTTGCTGGCTACATCATCTCGCTTCCGTATGATGAAT 1317

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              |||||     | || | | | ||  ||||  |||| ||  ||||  |||||  |  |||||
Db       1318 CTCCAGGGAGAGGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAATTCTGGAGTGTAC 1377

Qy       1075 CCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTTGCTG 1131
                |  |  |  |    |||| | |  ||  ||    |  |  |  |  | |||
Db       1378 ACGTTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACCGTGTCCTG 1437

Qy       1132 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1191
               ||    | || ||  |||| |      |||| ||||||   ||| |  ||   ||||||
```

```
Db       1438 GACAAGATCACAGACACCTTGATCCACCTGATGGCCAAAGCGGGGCCTGACTCTGCAGCAG 1497

Qy       1192 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1251
             ||   |  ||| || ||  | |||||  |  ||||||| ||| |||||||   |||||||
Db       1498 CAGCACCGGCGTCTCGCGCAGCTCCTCCTCATCCTGTCTCACTTCAGGCACATGAGTAAC 1557

Qy       1252 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1311
             || |||||||| |||||| ||||||||||||||| || ||||| ||  | |||||||||
Db       1558 AAAGGCATGGAGCATCTGTACAACATGAACTGCAAGAACGTGGTGCCCCTCTATGACCTG 1617

Qy       1312 CTGCTGGAGATGCTGAATGCCCACGTGCTTCGCG 1345
             ||||||||||||||| | ||||||   || | ||
Db       1618 CTGCTGGAGATGCTGGACGCCCACCGCCTGCACG 1651


RESULT 15
US-10-198-846-11305
; Sequence 11305, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MPI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11305
;   LENGTH: 6608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 6608
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-11305

  Query Match             26.9%;  Score 385.2;  DB 9;  Length 6608;
  Best Local Similarity   62.1%;  Pred. No. 2.5e-108;
  Matches  698;  Conservative    0;  Mismatches  383;  Indels   43;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || |||  || || || || || || || |||||||||
Db        898 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 957

Qy        331 TGGTCGTGTCAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        958 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAACTATTCAAGGACATAACGACTAT 1017

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGCCGCAAGAGCTGCCAGGGC 450
              || |||||||| || || ||||| || || |||||||||| || | ||||||||||||||
Db        1018 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGGC 1077

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||    || | |||  ||
Db        1078 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGCATACGAAAAGACCCA 1137

Qy        511 TGTGGGTACCGCCTTGTG----------------CGGAGACAGAGAAGTGCCGACGAGCA 554
               | ||     |  |  ||                    ||    | | | |  || | |
Db        1138 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGTACTTGCTACCCGAA 1197

Qy        555 GCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGGCGCCACGGCCCCCACTGCGCGAGCT 614
                  |    | |     | || |||  || |  ||    |||  |   ||    | |
Db        1198 TTCCCATCCATCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGAT 1257

Qy        615 ----------------------GCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCAC 654
                                     ||| |      |||||  |||| ||| | ||    |
Db        1258 CAAACGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGC 1317

Qy         654 CCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATC---AGCCGCCCCAGTGCGCCCTCAC 710
               | |  |||| |||||||| ||  |   |   |          || |     |||||||
Db        1318 CTTGTTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAG 1377

Qy         711 CGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGAT 770
               || || || |||||   |  ||||||||  |||| |||| |||| |||| ||||||||
Db        1378 TGAAGCTTCGATGATGGGCTTACTGACCAACCTGGCAGACAGGGAGCTGGTTCACATGAT 1437

Qy         771 CAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCT 830
               || |||||| |||| | | || ||||||||||  | | |||    || || || | ||
Db        1438 CAACTGGGCGAAGAGGGTGCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCT 1497

Qy         831 CTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCA 890
               | ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |
Db        1498 TGTAGAATGTGCCTGGCTAGACATCCGGATCATTGGTCTGGTGTGGGCTCCATGGAGCA 1557

Qy         891 CCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGT 950
               ||| |  |||||  | |||||||| |  |  |  |||||||| |  |||| |||| ||
Db        1558 CCCAGTGAAGCTACTGTTTGCTCCTAACTTGCTGTTGGACAGGAACCAGGGAAAATGTGT 1617

Qy         951 AGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTT 1010
               ||| || ||  |||| ||||| |||||||| |||| ||  | ||  ||||  ||   | |
Db        1618 AGAGGGCATGGTGGAGATCTTCGACATGCTGCTGGCTACATCATCTCGGTTCCGCATGAT 1677

Qy        1011 AAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTAT 1070
               || || ||      | || | | | ||  |||  | || ||  |||| ||||  | |
Db        1678 GAATCTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATTATTTGCTTAATTCTGGAGT 1737

Qy        1071 CTACCCTCTGGTCACAGCGACCCAGGATGCT---GACAGCAGCCGGAAGCTGGCTCACTT 1127
               |||| |  |  |  |  |||| | |  ||   ||    |  |  | |   |  |  |
Db        1738 CTACACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACCATATCCACGAGT 1797

Qy        1128 GCTGAACGCCGTGACCGATGCTTTGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCA 1187
                ||| ||    | || ||  |||| |      |||| ||||||   ||| |  ||   ||
Db        1798 CCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCA 1857

Qy        1188 GCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               |||||  |  ||| |||| | |||| |  |||| |||||| |||||||   |||
Db        1858 GCAGCAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAG 1917

Qy        1248 TAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGA 1307
               |||||| |||||||| ||||||  || ||||||||||||| || |||||  |  | |||||
Db        1918 TAACAAAGGCATGGAGCATCTGTACAGCATGAAGTGCAAGAACGTGGTGCCCCTCTATGA 1977

Qy        1308 CCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCA 1351
               |||||||||||||||||| | ||||||   || |  | |  ||
Db        1978 CCTGCTGCTGGAGATGCTGGACGCCCACCGCCTACATGCGCCCA 2021


RESULT 16
US-10-052-092-23
; Sequence 23, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: Patentin version 3.1
; SEQ ID NO 23
;   LENGTH: 1978
```

```
;   TYPE: DNA
;   ORGANISM: Hamster
US 10 052 092 23

  Query Match           25.5%;  Score 366.2;  DB 9;  Length 1978;
  Best Local Similarity   60.7%;  Pred. No. 1e 102;
  Matches  674;  Conservative    0;  Mismatches  498;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || ||||| || || || || || ||||| |||
Db        713 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTACGCCTCTGGCTACCATTATGGGGTC 772

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| |||||||| || |||| || || || ||||| |||||||||| ||||| ||
Db        773 TGGTCCTGTGAAGGCTGCAAGGCTTTCTTCAAGAGAAGTATTCAAGGACACAATGACTAC 832

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCC 450
              || |||||||||||||| || || |||||||| || ||| || | ||||||||||||||||
Db        833 ATGTGTCCAGCTACAAACCAATGCACAATCGACAACAACAGGAGAAAGAGCTGCCAGGCC 892

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTCAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||| | || || ||||||||||||||| || ||| |||| |||| |||| ||| | |
Db        893 TGCAGGCTGCGCAAGTGTTACCAACTAGGCATGATCAAACGTCCCATACGGCAAACACCCC 952

Qy        511 TGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGC 570
              | ||      | | ||        || ||      ||||   |      | |   || |
Db        953 AGAGGAGGGAGAATGCTGAAACACAAGCGCCAGAGAGACGACTTGGAAGGCAGGAACGAC 1012

Qy        571 AAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGG--- --- - 622
              | ||       | | | |      || ||    |  ||       |  || |
Db        1013 ATGGGGCCTTCAGGAGACATGAGGGCCACCAACCTTTGGCCTAGTCCTCTTGTGATTAAG 1072

Qy        623              - -----        ACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                                          || |||   |||| ||  | ||     || |
Db        1073 CACACTAAGAAGAACAGCCCTGCCTTGTCCTTGACAGCCGACCAAATGGTCAGTGCCTTG 1132

Qy        658 CTGGAGGCTGAGCCGCCCCATGT---GCTGATCAGCCGCCCCACTGCGCCCTTCACCGAG 714
              |||| |||| ||||||    |                 ||   |   || |||| |||
Db        1133 TTGGATGCTGAACCGCCCTTAATCTATTCTGAATATGATCCTTCTAGACCTTTCAGCGAA 1192

Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGC 774
              || || |||||   |   |||||||  |||| |||| ||||||||| || ||||||| |
Db        1193 GCTTCGATGATGGGATTATTGACCAACCTGGCAGACAGGGAGTTGGTTCATATGATCAAC 1252

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTG 834
              |||: |||   | || ||||||  || | |  ||    || || || |  ||| ||
Db        1253 TGGGCAAAGAGATGCCAGGCTTTGGAGACTTAAATCTCCATGATCAGGTCCACCTCCTG 1312

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||| |    ||| |||||  |  | ||||| || || || ||||||||  || || |||||
Db        1313 GAGTGTGCCTGGTTGGAGATCCTGATGATTGGTCTCATCTGGCGCTCCATGGAACACCCA 1372

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAA 954
              || |||||  | |||||||  || |  | |||||||| || |||| || || || ||
Db        1373 GGGAAGCTCCTGTTTGCTCCTAATTTGCTCCTGGACAGGAATCAGGGCAAGTGTGTGGAG 1432

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| |||||||||||| | |||| ||  |  |  |||| ||  | | |
Db        1433 GGCATGGTGGAGATCTTGACATGTTGCTGGCTACATCAGCTCGGTTCCGCATGATGGAC 1492

Qy        1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
               || ||     | || | | | ||  |||| | || ||  |||| ||||  |  |||||
Db        1493 CTGCAGGGAGAGGAGTTTGTGTGCCTCAAATCTATCATTTTGCTTAATTCTGGAGTGTAC 1552

Qy        1075 CC   TCTGGTCACAGCGACCCACGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTG 1134
               |   |||| ||  |  |    || |  ||    |  |  |  | |||
Db        1553 ACATTTCTGTCCAGCACCTTGAAGTCTCTGGAGGAGAAGGACCACATCCACCGGGTCCTG 1612

Qy        1135 AACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAG 1194
               |     | || ||  |||| ||    |||| ||||     ||| |  |    ||||||
Db        1613 GATAAGATCACAGACACTTTGATTCACCTGATGGCCAAAGCTGGCCTGACACTGCAGCAG 1672

Qy        1195 CAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAAC 1254
               ||       || |||||  | |||||  |  | || ||||||  || ||||   |||||||
Db        1673 CAGCATCGTCGTCTGGCCCAGCTCCTCCTCATTCTTTCCCACATCCGGCACATGAGTAAC 1732

Qy        1255 AAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTG 1314
               || |||||||| || ||    |||||||| |||| || || || || ||   | |||||||
Db        1733 AAAGGCATGGAGCACCTCTACAACATGAAATGCAAGAACGTTGTACCCTTCTATGACCTG 1792

Qy        1315 CTGCTGGAGATGCTGAATGCCCACGTGCTTC 1345
                |||||||||||  || |||| |||   || |
Db        1793 TTGCTGGAGATGTTGGATGCTCACCGCCTGC 1823


RESULT 17
US-10-052-092-5
; Sequence 5, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Bopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 2106
;   TYPE: DNA
;   ORGANISM: Human
US 10-052-092-5

  Query Match           22.8%;  Score 326.6;  DB 9;  Length 2106;
  Best Local Similarity   60.8%;  Pred. No. 2e-90;
  Matches  624;  Conservative    0;  Mismatches  364;  Indels   39;  Gaps    4;

Qy        271 AGGGATGCTCACTTCTGCGGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || ||||| || || || || || ||||| |||
Db        538 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 597

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| |||||| || || |||||||| || || ||||| |||||||||||||| || |||
Db        598 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 657

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || |||||||||| || | ||||||||||||||
Db        658 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 717

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || |||||||||||||||| |||| ||||    || | |||  ||
Db        718 TGCCGGCTCCGCAAATGCTACGAAGTGGCAATGATGAAAGGTGGGATACGAAAAGACCGA 777

Qy        511 TGTGG-----------GTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTG 558
              | ||           | | | |     | |||| |     | | | | |
Db        778 AGAGGAGGGAGAATGTTGAAACACAAGCGCCAGAGAGATGATGGGGAGGGCAGGGGTGAA 837

Qy        559 CACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGC- --------- 607
                | | |      | || |||  || |   |||  |  ||
Db        838 GTGGGGTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCAAA 897

Qy        608 ----------GGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTC 657
                       | |    |||      ||||||   |||| ||| | ||     || |
Db        898 CGCTCTAAGAAGAACAGCCTGGCCTTGTCCCTGACGGCCGACCAGATGGTCAGTGCCTTG 957

Qy        658 CTGGAGGCTGAGCCGCCCCATGTGCTGATC- -AGCCGCCCCACTGCGCCCTTCACCGAG 714
              |||| |||||||| |||    |      |          || |     |||||| ||
Db        958 TTGGATGCTGAGCCCCCCATACTCTATTCCGAGTATGATCCTACCAGACCCTTCAGTGAA 1017
```

```
Qy        715 GCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGCAGTTGCTACACATCATCAAC 774
              || || ||||||   |  ||||||||  |||| |||| |||| |||| ||||||||||  |
Db       1018 GCTTCGATGATGGCCTTACTGACCAACCTGGCAGACAGGGAGCTGCTTCACATGATCAAC 1077

Qy        775 TGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGGCGCTGCTGC 834
              ||||| |||| | | || ||||||||||  | | |||    || || || |  ||  |
Db       1078 TGGGCGAAGAGGGTGCCAGGCTTTGTGGATTTGACCCTCCATGATCAGGTCCACCTTCTA 1137

Qy        835 GAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCC 894
              ||  |    ||| | ||| |  | ||||| || ||  | |||||||| || || |||||
Db       1138 GAATGTGCCTGGCTAGAGATCCTGATGATTGGTCTGCTGGGCTCCATGGAGCACCCA 1197

Qy        895 GGCAAGCTCATCTTTGCTCCAGATCTTCTTCTGGACAGGGATGAGGGCAAATGCCTACAA 954
              |  ||||| | |||||||  |  |  |  |||||||| |  |||| |||| ||||
Db       1198 GTGAAGCTACTGTTTGCTCCTAACTTGCTCTTGGACAGGAACCAGGGAAAAATGTGTACAG 1257

Qy        955 GGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAA 1014
              || ||  |||| |||||  |||||||| ||||| ||  | ||  |||| ||    | | ||
Db       1258 GGCATGGTGGACATCTTCGACATGCTGCTGGCTACATCATCTCCGTTGCGCCATGATGAAT 1317

Qy       1015 CTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTAC 1074
              || ||     | || | | | ||  ||||  | || ||  |||| |||||  |  |||||
Db       1318 CTGCAGGCAGAGGAGTTTGTGTGCCTCAAATCTATTATTTGCTTAATTCTGCACTGTAC 1377

Qy       1075 CCTCTGGTCACAGGACCCAGCATGCT---GACAGCACCCCGAAGCTGCGTCACTTGCTC 1131
              |  |  | |    |||| | |  ||   ||     |  |    |  |   | |||
Db       1378 ACATTTCTGTCCAGCACCCTGAAGTCTCTGGAAGAGAAGGACATATCCAGGAGTGCTC 1437

Qy       1132 AACGCCGTGACCGATGCTTTGTTTGGGCATTGCCAAGAGCGGCATCTCCTCCCACAGAG 1191
              ||    | || ||  |||| |     |||| ||||||   ||| |  ||   ||||||
Db       1438 GACAAGATCACAGACACTTTGATCCACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAG 1497

Qy       1192 CAATCCATCGCCCTGGCTAAGCTCCTGATGTCCTTCGCCACCTCAGCTATGTTAAAC 1251
              ||   |  ||| |||||  | |||||  |  |||| |||||| |||||||   ||| |||
Db       1498 CAGCACCAGCGGCTGGCCCAGCTCCTCCTCATCCTCTCCCACATCAGGCACATGAGGAAC 1557

Qy       1252 AAGGGCA 1258
              |||| |
Db       1558 CAGGGAA 1564


RESULT 18
US-10-259-864-3
; Sequence 3, Application US/10259864
; Publication No. US20030077645A1
; GENERAL INFORMATION:
;  APPLICANT: Hager, Gordon L
;  TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfamil
;  TITLE OF INVENTION:  Receptor Ligands, and Methods and Kits For Detection and Charact
;  TITLE OF INVENTION:  Ligands
;  FILE REFERENCE: 24799
;  CURRENT APPLICATION NUMBER: US/10/259,864
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: US 60/325,178
;  PRIOR FILING DATE: 2001-09-28
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;  LENGTH: 3300
;  TYPE: DNA
;  ORGANISM: Rat/human Chimera
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(3300)
;  OTHER INFORMATION:
US-10-259-864-3

  Query Match             21.4%;  Score 306.8;  DB 9;  Length 3300;
  Best Local Similarity   57.9%;  Pred. No. 3.4e-84;
  Matches  645;  Conservative    0;  Mismatches  432;  Indels   29;  Caps    4;

Qy        284 TCTGCGCTGTCTGCAGGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 343
              |||||   || |||  |||| | || || |||| ||| || || ||   | | |||| |

Db       2075 TCTGCGTGGGGGGCTCCGATGAAGCTTCAGCATGTCATTACGGGCTCCTGACATCTGGAA 2134

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | || || |  || |||||||||     |  ||||||  |  ||||  |||| | |  |
Db       2135 GCTGCAAAGTATTCTTTAAAAGAGCAGTGGAAGCACAGCACAATTACCTTTGTGCTGGAA 2194

Qy        404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||  | || |  || |||||||  ||  | || | |||||  || |||||   |||||
Db       2195 GAAACGATTGCATCATTCATAAAATTCCAAGGAAAAACTGCCCAGCATGCCGCTATCGGA 2254

Qy        464 AGTCTTACGAAGTGGGAAGTGTGAAGTGTGGCTCCCGGAGAGAGACATGTGGGTACCGCC 523
              | |||    | |  ||||||        |         | | ||| |       |  |
Db       2255 AATGTCTTCAGGCTGGAATGAACCTTGAAGCTCGAAAAACAAAGAAAAAAATCAAAGGA 2314

Qy        524 TTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAA 583
              ||  ||  || ||  | ||       |     | |     | |      | |    || |  |
Db       2315 TTCAGC-AAGCCAGTGCAGGAGTCTCACAAGACACTTCGGAAAATCCTAACAAAACAATA 2373

Qy        584 GTGCCGGCCACGGCCCCCGAGTGCG----------GCAGCTGCTGCTGGACGCCCTGAGC 633
              ||  | ||  |    ||  ||  |              || |   |||||||| |    |
Db       2374 GTTCCTCCACGATTACCACAGCTCACCCCTACCTTGGTGTCACTGCTGGACGTGATTGAA 2433

Qy        634 CCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGC 693
              ||||||  | |              |||   |||    || ||     ||  | |
Db       2434 CCCGACGTCTTCTATGCAGCATATGATAGCTCTGTTTCCGCAGGTCAGCCCCCATACTC 2493

Qy        694 ----------------GCCAGTGCGGCCGTTCACCGAGGCCTCCATGATGATGTCCCTGACC 738
                             || |     ||||||| ||  || || || ||||||   |  ||||||
Db       2494 TATTCCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGCCTTACTGACC 2553

Qy        739 AAGTTGGCCGACAAGCAGTTGCTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTT 798
              ||  |||| |||| |||| |||| ||||||||| ||||||  |||| | | | ||||||
Db       2554 AACCTGGCAGACAGGGAGCTGCTTCACATGATCAACTGGGCGAAGAAGATTCCCGGCTTT 2613

Qy        799 GTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTA 858
              |||||  | | |||    || || || |  ||  | ||  |    ||| | ||| |  |
Db       2614 GTGGATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTG 2673

Qy        859 ATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGAT 918
              ||||| || ||  | |||||||| || || ||||| |  ||||  | |||||||  |
Db       2674 ATGATTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAAC 2733

Qy        919 CTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAAATCTTTGACATG 978
              |  |  |||||||| |  |||| |||| |||| || ||  |||| |||| ||||||
Db       2734 TTGCTCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATG 2793

Qy        979 CTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGT 1038
              || |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||
Db       2794 CTGCTGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGC 2853

Qy       1039 GTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGGGACCCAGGAT 1098
              ||||  | || ||  |||| ||||  |  |||| |  |  |  |    |||| | |
Db       2854 CTCAAATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAG 2913

Qy       1099 GCT---GACAGCAGCCCGAAGCTGGCTCACTTGCTGAACGCCCTGACCCATGCTTTGGTT 1155
              ||   ||    |      |  |     |  | ||| ||    | || ||  |||| |
Db       2914 TCTCTGCAAGAGAAGCACCATATCCACCGACTCCTGGACAAGATCACAGACACTTTGATC 2973

Qy       1156 TGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCCCCTGGCTAACCTC 1215
                  |||| ||||||   ||| |  ||   ||||||||   |  ||| |||||  | |||
Db       2974 CACCTGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTC 3033

Qy       1216 CTGATGCTCCTGTCCCACATCAGGCATCCGAGTAACAAGGGCATGCAACATCTGCTCAAC 1275
              ||  |  |||| |||||| |||||||   |||||||| |||||||| ||||||  || |
Db       3034 CTCCTCATCCICTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGC 3093

Qy       1276 ATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCAC 1335
              |||||||||| || |||| ||  | ||||||||||||||||||||||||| | ||||||
Db       3094 ATGAAGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCAC 3153

Qy       1336 GTGCTTCGCGGGTGCA 1351
                 || |  | |  ||
Db       3154 CGCCTACATGCGCCCA 3169
```

```
RESULT 19
US-09-864-761-16488
; Sequence 16488, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Ranzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Acomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: GB 24264.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16488
;  LENGTH: 586
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AL161756.1
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL   0.46
US-09-864-761-16488

  Query Match             21.1%;  Score 402;  DB 10;  Length 586;
  Best Local Similarity   100.0%;  Prod. No. 4.1e-83;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         494 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          33 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAG  92

Qy         554 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCCGAGTGCGGGAG 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          93 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCCGAGTGCGGGAG 152

Qy         614 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         153 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 212

Qy         674 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         213 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 272

Qy         734 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         273 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 332

Qy         794 GG 795
               ||
Db         333 GG 334


RESULT 20
US-09-864-761-32872
; Sequence 32872, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Ranzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Acomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: GB 24264.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32872
;  LENGTH: 400
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.46
;   OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 6.00e-52
;   OTHER INFORMATION: EST_HUMAN HIT: BE747838.1, EVALUE 6.90e-02
;   OTHER INFORMATION: NT HIT: AB006590.1  EVALUE 0.00e+00
US-09-864-761-32872

  Query Match             20.9%;  Score 300;  DB 10;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-82;
  Matches  300;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        494 GCTCCCCGGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCCACGAGC 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTCCCCGGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCCACGAGC 60

Qy        554 AGCTGCACTGTGCCCGCAAGCCCAAGAGAAGTGGCCGCCACGCGCCCCGAGTGCGGGAGC 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCTGCACTGTGCCCGCAAGCCCAAGAGAAGTGGCCGCCACGCGCCCCGAGTGCGGGAGC 120

Qy        614 TGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGC 180

Qy        674 CCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACGGAGGCCTCCATGATGATGTCCCT 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACGGAGGCCTCCATGATGATGTCCCT 240

Qy        734 TGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCG 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCG 300


RESULT 21
US-10-052-092-2
; Sequence 2, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 1644
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Estrogen Receptor Ligand Binding Domain Fusion
US-10-052-092-2

  Query Match             19.9%;  Score 285.6;  DB 9;  Length 1644;
  Best Local Similarity   63.8%;  Pred. No. 8.8e-78;
  Matches  467;  Conservative    0;  Mismatches  259;  Indels    6;  Gaps    2;

Qy        626 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 685
              ||||||   |||| ||| | | ||     || |  |||| |||||||| |||    |
Db        782 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGCTGGATGCTGAGCCCCCCATACCCTATT 841

Qy        686 TC---AGCCGCCCCAGTGCGCCCTTCACGGAGGCCTCCATGATGATGTCCCTGACCAACCT 742
               |         || |     |||||||  || || || ||||||   |  ||||||||
Db        842 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 901

Qy        743 TGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGC 802
              |||| |||| |||| |||| ||||||||| |||||| |||| | | || |||||||||
Db        902 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 961

Qy        803 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGCAGAGCTGTTGGATGCAGTGTTAATGA 862
              | | | |||    || || || | ||  | ||  |    ||| | ||| |  | ||||
Db        962 ATTTCACCCTCCATCATCAGTCCACTTCTAGAATGTCCCTGGCTAGACATCCTGATCA 1021

Qy        863 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 922
              | || ||  | |||||||| || || ||||| |  |||||  | ||||||||  | |
Db       1022 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1081

Qy        923 TTCTCGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 982
              |  |||||||| |  |||| ||||| ||||| || ||  |||| ||||| ||||||||  |
Db       1082 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1141

Qy        983 TGGCAACTACTTCAAGCTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1042
              |||| ||  | ||  |||| ||   | | || || ||      | || | | | ||  |||
Db       1142 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1201

Qy       1043 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1101
              |  | ||  ||  |||| |||||  |  ||||| |  |  |  |  |    |||| | |  ||
Db       1202 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1261

Qy       1102 --GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1159
                ||    |    |  |    | ||| ||    | || ||  ||||| |
Db       1262 TGGAAGAGAAGGACCATATCCACCGAGTCCTGCACAACATCACAGACACTTTGATCCACC 1321

Qy       1160 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1219
              |||| ||||||   ||| |  ||   ||||||||   |  ||| |||||  | |||||
Db       1322 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1381

Qy       1220 TGCTCCTGTCCCACCTCAGCCATCGCACTAACAAGCCCATGCAACATCTGCTCAACATGA 1279
              |  |||| |||||| |||||||   |||||||| |||||||| ||||||  || ||||
Db       1382 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1441

Qy       1280 AGTGCAAAAATGTGGTGCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1339
              ||||||| || |||| ||  | ||||||||||||||||||||||| | ||||||   |
Db       1442 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCC 1501

Qy       1340 TTCGCGGGTGCA 1351
              | |  | |  ||
Db       1502 TACATGCGCCCA 1513


RESULT 22
US-10-052-092-19
; Sequence 19, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therape
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
;   LENGTH: 11522
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Vector pER8
US-10-052-092-19
```

```
  Query Match             19.9%;  Score 285.6;  DB 9;  Length 11522;
  Best Local Similarity   63.8%;  Pred. No. 2.6e-77;
  Matches  467;  Conservative    0;  Mismatches  259;  Indels    6;  Gaps    2;

Qy        626 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 685
              ||||||   |||| ||| | ||     || |  |||| ||||||||| |||    |
Db        880 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 939

Qy        686 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 742
              |         || |     |||||||  || || || ||||||   |  ||||||||
Db        940 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 999

Qy        743 TGGCCGACAAGGAGTTGGTGACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 802
              |||| |||| |||| |||| |||||||||| |||||| ||| | | || ||||||||||
Db       1000 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1059

Qy        803 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 862
              | | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1060 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1119

Qy        863 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 922
              | || ||  | |||||||| || || |||| |   ||||  | ||||||||  |  |
Db       1120 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1179

Qy        923 TTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 982
              |  |||||||| |  |||| |||| |||| || ||  |||| |||| |||||||| |
Db       1180 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1239

Qy        983 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1042
              |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||  |||
Db       1240 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1299

Qy       1043 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT  1101
              |  | || ||  |||| ||||  |  |||| |  |  | |  |    |||| | |  ||
Db       1300 AATCTATTATTTTGCTTAATTCTGGAGTGTACATTTCTGTCCAGCACCCTGAAGTCTC 1359

Qy       1102   GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1159
                ||    |      |  |    |  ||| ||    | || ||  |||| |
Db       1360 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1419

Qy       1160 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1219
              |||| ||||||   ||| |  ||   |||||||  |  ||| |||| | ||||
Db       1420 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1479

Qy       1220 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1279
              |  |||| |||||| |||||||   |||||||| |||||||| |||||  || |||||
Db       1480 TCATCCTCTCCCACATCAGGCACATGAGTAACAAACGCATGGAGCATCTGTACAGCATGA 1539

Qy       1280 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1339
              ||||||| || ||||| ||  | ||||||||||||||||||||||| | ||||||   |
Db       1540 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGACATGCTGGACGCCCACCGCC 1599

Qy       1340 TTCGCGGGTGCA 1351
              | |  | |  ||
Db       1600 TACATGCGCCCA 1611


RESULT 23
US-09-853-033-3
; Sequence 3, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;  LENGTH: 1983
;  TYPE: DNA
;  ORGANISM: Artificial sequence
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(1983)
;  OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;  OTHER INFORMATION: Homosapiens-Bacteriophage P1
US-09-853-033-3

  Query Match             19.8%;  Score 284;  DB 10;  Length 1983;
  Best Local Similarity   63.7%;  Pred. No. 3e-77;
  Matches  466;  Conservative    0;  Mismatches  260;  Indels    6;  Gaps    2;

Qy        626 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 685
              ||||||   |||| ||| | ||     || |  |||| ||||||||| |||    |
Db       1121 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1180

Qy        686 TC---AGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 742
              |         || |     |||||||  || || || ||||||   |  ||||||||
Db       1181 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1240

Qy        743 TGGCCGACAAGGAGTTGGTGACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 802
              |||| |||| |||| |||| |||||||||| |||||| ||| | | || ||||||||||
Db       1241 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1300

Qy        803 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 862
              | | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1301 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1360

Qy        863 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 922
              | || ||  | |||||||| || || |||| |   ||||  | ||||||||  |  |
Db       1361 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGGGAAGCTACTGTTTGCTCCTAACTTGC 1420

Qy        923 TTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 982
              |  |||||||| |  |||| |||| |||| || ||  |||| |||| |||||||| |
Db       1421 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1480

Qy        983 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1042
              |||| ||  | ||  |||| ||   | | || || ||     | || | | | ||  |||
Db       1481 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1540

Qy       1043 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT  1101
              |  | || ||  |||| ||||  |  |||| |  |  | |  |    |||| | |  ||
Db       1541 AATCTATTATTTTGCTTAATTCTGGAGTGTACATTTCTGTCCAGCACCCTGAAGTCTC 1600

Qy       1102 --GACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1159
                ||    |      |  |    |  ||| ||    | || ||  |||| |
Db       1601 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1660

Qy       1160 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1219
              |||| ||||||   ||| |  ||   |||||||  |  ||| |||| | ||||
Db       1661 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1720

Qy       1220 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1279
              |  |||| |||||| |||||||   |||||||| |||||  || |||||
Db       1721 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAAGAATGGAGCATCTGTACAGCATGA 1780

Qy       1280 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1339
              ||||||| || ||||| ||  | ||||||||||||||||||||||| | ||||||   |
Db       1781 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCCACCGCC 1840

Qy       1340 TTCGCGGGTGCA 1351
              | |  | |  ||
Db       1841 TACATGCGCCCA 1852


RESULT 24
US-09-853-033-7
; Sequence 7, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;    LENGTH: 1983
;    TYPE: DNA
;    ORGANISM: Artificial sequence
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1)..(1983)
;    OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;    OTHER INFORMATION: Homosapiens-Bacteriophage P1
US-09-853-033-7

  Query Match             19.6%;  Score 280.8;  DB 10;  Length 1983;
  Best Local Similarity   63.4%;  Pred. No. 3e-76;
  Matches  464;  Conservative    0;  Mismatches  262;  Indels    6;  Gaps    2;

Qy        626 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 685
              ||||||   |||| ||| | ||     || |  |||| ||||||||| |||    |
Db       1121 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1180

Qy        686 TC---AGCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 742
               |         || |     ||||||| || || || ||||||   |  ||||||||
Db       1181 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1240

Qy        743 TGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 802
              |||| |||| |||| |||| |||||||||| |||||| |||| | | || ||||||||||
Db       1241 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1300

Qy        803 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 862
              |  | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1301 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1360

Qy        863 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCACATCTTG 922
              | || ||  | |||||||| || || ||||| || ||||  | |||||||  | |  |
Db       1361 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGGGAAGCTACTGTTTGCTCCTAACTTGC 1420

Qy        923 TTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 982
              |  |||||||| |  |||| |||| |||| || ||  |||| |||| ||||||||  |
Db       1421 TCTTGGACAGGAACCAGGGAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1480

Qy        983 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1042
              |||| ||  | ||  |||| ||    | | || || ||      | || | | | ||  |||
Db       1481 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1540

Qy       1043 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1101
              |  | ||  ||  |||| ||||  |  |||| |  |  |  |    |||| | |  ||
Db       1541 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1600

Qy       1102 --GACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGCTTTGGG 1159
                ||   |    |  |    |  ||| ||    | || ||  ||||| |
Db       1601 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1660

Qy       1160 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1219
              |||| ||||||   ||| |  ||   ||||||||   |  ||| ||||  | |||||
Db       1661 TGATGGCCAAGGCAGGCCTGACCTTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1720

Qy       1220 TGCTCCTGTCCCACGTCAGGCATGCGACTAACAAGGGCATGGAACATCTGCTCAACATGA 1279
              |  |||| |||||| |||||||   |||||||| |||||||| |||||| || |||||
Db       1721 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1780

Qy       1280 AGTGCAAAAAAGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1339
              ||||||| || ||||| ||  | ||||||||||||||||||  |  | | ||||||   |
Db       1781 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGGCGGGGGACGCCCACCGCC 1840

Qy       1340 TTCGCGGGTGCA 1351
              | |  | |  ||
Db       1841 TACATGCGCCCA 1852


RESULT 25
US-09-853-033-5
; Sequence 5, Application US/09853033
; Patent No. US20020100068A1
; GENERAL INFORMATION:
;  APPLICANT: CHAMBON, PIERRE
;  APPLICANT:  METZGER, DANIEL
;  TITLE OF INVENTION: TRANSGENIC MOUSE FOR TARGETED RECOMBINATION
;  TITLE OF INVENTION:  MEDIATED BY MODIFIED CRE-ER
;  FILE REFERENCE: 065691/0222
;  CURRENT APPLICATION NUMBER: US/09/853,033
;  CURRENT FILING DATE:  2001-05-11
;  PRIOR APPLICATION NUMBER: FR 00/12570
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;    LENGTH: 1983
;    TYPE: DNA
;    ORGANISM: Artificial sequence
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1)..(1983)
;    OTHER INFORMATION: Description of Artificial Sequence: Chimeric sequence
;    OTHER INFORMATION: Homosapiens-Bacteriophage P1
US-09-853-033-5

  Query Match             19.5%;  Score 279.2;  DB 10;  Length 1983;
  Best Local Similarity   63.3%;  Pred. No. 9.4e-76;
  Matches  463;  Conservative    0;  Mismatches  263;  Indels    6;  Gaps    2;

Qy        626 CCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGA 685
              ||||||   |||| ||| | ||     || |  |||| ||||||||| |||    |
Db       1121 CCCTGACGGCCGACCAGATGGTCAGTGCCTTGTTGGATGCTGAGCCCCCCATACTCTATT 1180

Qy        686 TC---AGCCGCCCCAGTGCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGT 742
               |         || |     ||||||| || || || ||||||   |  ||||||||
Db       1181 CCGAGTATGATCCTACCAGACCCTTCAGTGAAGCTTCGATGATGGGCTTACTGACCAACC 1240

Qy        743 TGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGG 802
              |||| |||| |||| |||| |||||||||| |||||| |||| | | || ||||||||||
Db       1241 TGGCAGACAGGGAGCTGGTTCACATGATCAACTGGGCGAAGAGGGTGCCAGGCTTTGTGG 1300

Qy        803 AGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGA 862
              |  | | |||    || || || |  ||  | ||  |    ||| | ||| |  | ||||
Db       1301 ATTTGACCCTCCATGATCAGGTCCACCTTCTAGAATGTGCCTGGCTAGAGATCCTGATGA 1360

Qy        863 TGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTG 922
              | || ||  | |||||||| || || ||||| |  |||||  | ||||||||  |  |
Db       1361 TTGGTCTCGTCTGGCGCTCCATGGAGCACCCAGTGAAGCTACTGTTTGCTCCTAACTTGC 1420

Qy        923 TTCTGGACACGGATGACGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCC 982
              |  |||||||| |  |||| |||| |||| || ||  |||| |||| ||||||||  |
Db       1421 TCTTGGACAGGAACCAGGCAAAATGTGTAGAGGGCATGGTGGAGATCTTCGACATGCTGC 1480

Qy        983 TGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCA 1042
              |||| ||  | ||  |||| ||    | | || || ||      | || | | | ||  |||
Db       1481 TGGCTACATCATCTCGGTTCCGCATGATGAATCTGCAGGGAGAGGAGTTTGTGTGCCTCA 1540

Qy       1043 AGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACCCAGGATGCT- 1101
              |  | ||  ||  |||| ||||  |  |||| |  |  |  |    |||| | |  ||
Db       1541 AATCTATTATTTTGCTTAATTCTGGAGTGTACACATTTCTGTCCAGCACCCTGAAGTCTC 1600

Qy       1102 --GACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGG 1159
                ||   |    |  |    |  ||| ||    | || ||  ||||| |
Db       1601 TGGAAGAGAAGGACCATATCCACCGAGTCCTGGACAAGATCACAGACACTTTGATCCACC 1660
```

```
Qy       1160 TGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGA 1219
             |||| ||||||   ||| | ||   ||||||||   | ||| |||||  | |||||
Db       1661 TGATGGCCAAGGCAGGCCTGACCCTGCAGCAGCAGCACCAGCGGCTGGCCCAGCTCCTCC 1720

Qy       1220 TGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGA 1279
             |  |||| |||||| |||||||   |||||||| |||||||| ||||||  || |||||
Db       1721 TCATCCTCTCCCACATCAGGCACATGAGTAACAAAGGCATGGAGCATCTGTACAGCATGA 1780

Qy       1280 AGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGC 1339
             ||||||| || ||||| ||  | ||||||||||||||||||  |  | | ||||||  |
Db       1781 AGTGCAAGAACGTGGTGCCCCTCTATGACCTGCTGCTGGAGGCGGCGGACGCCCACCGCC 1840

Qy       1340 TTCGCGGGTGCA 1351
             | |  | |  ||
Db       1841 TACATGCGCCCA 1852


RESULT 26
US-09-918-995-15134
; Sequence 15134, Application US/09918995
; Publication No. US20030073624A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/245,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 48054
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15134
;  LENGTH: 451
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(451)
;  OTHER INFORMATION: n  A,T,C or G
US-09-918-995-15134

  Query Match             19.4%;  Score 278;  DB 9;  Length 451;
  Best Local Similarity   90.2%;  Pred. No. 9.7e-76;
  Matches  296;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        496 TCCCGGACAGAGAGATGTGGGTACCGCCTTGTGCCGGAGACAGAGAAGTGCCGACGAGCAG 555
              |  ||||||||||||||||||||||| ||||||| || ||||| |||||||||||| |||
Db         49 TNNAGGAGAGAGATGTGGGTACCGACTTGTGCTGAAACAGATAAGTGCCGACGATCAG 108

Qy        556 CTGCACTGTGCCGGCAAGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTG 615
              ||||||||||||||||.|||| |||||| ||||| ||| ||||||||||||||||||||
Db        109 CTGCACTGTGCCGGCAAGCCAGGAGAAGGGCGGCTCACACGCCCCGAGTGCGGGAGCTG 168

Qy        616 CTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCC 675
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||| ||
Db        169 CTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGTCC 228

Qy        676 CATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTG 735
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        229 CATGTGCTGATCAGCCGACCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTG 288

Qy        736 ACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGC 795
              |||||||||||||||||.|||| ||||||||||||||||||||||||||||||||||||
Db        289 ACCAAGTTGGCCGACAAGGAGTCGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGT 348

Qy        796 TTGTGGAGCTCAGCCTGTTCGACCAAG 823
                |||       | |    |||| ||
Db        349 GATGTACCCTCTGGGCACAGCGACCAG 376


RESULT 27
US-09-737-255-4
```

```
; Sequence 4, Application US/09737255
; Patent No. US20020073441A1
; GENERAL INFORMATION:
;  APPLICANT: Rehemtulla, Alnawaz
;  APPLICANT:  Ross, Brian D.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
;  TITLE OF INVENTION:  PROTEOLYTIC ACTIVITY
;  FILE REFERENCE: 11203-003001
;  CURRENT APPLICATION NUMBER: US/09/737,255
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;  LENGTH: 973
;  TYPE: DNA
;  ORGANISM: Mus musculus
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(973)
US-09-737-255-4

  Query Match             18.6%;  Score 267;  DB 10;  Length 973;
  Best Local Similarity   64.5%;  Pred. No. 3.8e-72;
  Matches  415;  Conservative    0;  Mismatches  225;  Indels    3;  Gaps    1;

Qy        703 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 762
              |||||||  || |||| ||||||   |   |||||||  | || || | |||| ||||
Db        184 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 243

Qy        763 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAA 822
              || ||||||| |||||| ||||   | || ||||||| |||  | |  ||    || ||
Db        244 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATGATCAG 303

Qy        823 GTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCA 882
              || |  ||  | ||| |    ||| ||||| |  | ||||| || ||  | ||||||||
Db        304 GTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 363

Qy        883 ATTGACCACCCGGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 942
              || || ||||| || |||||| | ||||||||  |  |  | |||||||||| || | ||
Db        364 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 423

Qy        943 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTT 1002
              ||||| || |||| ||  |||| ||||||||||| | || || ||  |     ||||
Db        424 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 483

Qy       1003 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1062
              ||   | | || || ||     |||| | | | ||  |||| |||| ||  |||| |||
Db        484 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 543

Qy       1063 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1119
              ||| |  |||| |   |||| ||  |   ||  |  ||  |    |    |  |  |
Db        544 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 603

Qy       1120 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1179
              |   | ||| ||    | || ||  |||| |     |||| |||||    ||| |
Db        604 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 663

Qy       1180 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1239
              |    ||||||||         |||| ||| | ||||  |  | || ||||  || ||
Db        664 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 723

Qy       1240 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1299
              ||   |||||||| |||||||| ||||   |||||||| |||| || |||| ||
Db        724 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTGGTACCC 783

Qy       1300 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1342
              | |||||||||| |||||||| || ||||||||   ||||
Db        784 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 826


RESULT 28
US-09-908-154B-39
; Sequence 39, Application US/09908154B
```

```
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beeril, Roger
;  APPLICANT:  Schopter, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
;   LENGTH: 2739
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-153B-39

  Query Match             18.5%;  Score 265.4;  DB 9;  Length 2739;
  Best Local Similarity   64.4%;  Pred. No. 2.1e-71;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    3;  Gaps    1;

Qy        703 CCCTTCACCGAGGCCTCCATGATGATGTCCCTCACCAAGTTCGCCCACAAGGAGTTCGTA 762
              ||||||| || ||||| ||||||   |   |||||||  | || || | |||| ||||
Db        748 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 807

Qy        763 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAA 822
              || |||||| |||||| |||   | || ||||||| |||  | |  ||    || ||
Db        808 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGAGTTGAATCTCCATGATCAG 867

Qy        823 GTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTCTTAATGATGGGCTGATGTCCCCCTCA 882
              || |  ||  | ||| |   ||| ||||| |  | ||||| || ||  | ||||||||
Db        868 GTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTGTCTGGCGCTCC 927

Qy        883 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 942
              || || ||||| || |||||| | ||||||||  |  |  | ||||||||| || | ||
Db        928 ATGGAACACCCGGGGAAGCTCGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 987

Qy        943 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTT 1002
              ||||| || ||||| ||  |||| ||||||||||||| | || || ||  |      ||||
Db        988 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTGGGTTC 1047

Qy       1003 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1062
              ||   | | || || ||      |||| | | | ||  ||||  |||| ||  |||| |||
Db       1048 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1107

Qy       1063 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1119
              ||| |  ||||| |   ||||  ||   |      ||  |   ||      |   | |
Db       1108 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1167

Qy       1120 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1179
                |   | ||| ||    | || ||  ||||| |     |||| |||||    ||| |
Db       1168 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1227

Qy       1180 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1239
                |    ||||||||        ||||| ||| | |||||  |  | || |||||  || ||
Db       1228 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCGG 1287

Qy       1240 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1299
              ||   |||||||| ||||||||  |||||   ||||||||| ||||| || || || ||
Db       1288 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1347

Qy       1300 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1342
                | ||||||||||| |||||||| || ||||||||   ||||
Db       1348 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 1390


RESULT 29
```

```
US-09-908-153B-44/c
; Sequence 44, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beeril, Roger
;  APPLICANT:  Schopter, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 2739
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-153B-44

  Query Match             18.5%;  Score 265.4;  DB 9;  Length 2739;
  Best Local Similarity   64.4%;  Pred. No. 2.1e-71;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    3;  Gaps    1;

Qy        703 CCCTTCACCGAGGCCTCCATGATGATGTCCCTCACCAAGTTCGCCCACAAGGAGTTCGTA 762
              ||||||| || ||||| ||||||   |   |||||||  | || || | |||| ||||
Db       1932 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 1933

Qy        763 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAA 822
              || |||||| |||||| |||   | || ||||||| |||  | |  ||    || ||
Db       1932 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGAGTTGAATCTCCATGATCAG 1873

Qy        823 GTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTCTTAATGATGGGCTGATGTCCCCCTCA 882
              || |  ||  | ||| |   ||| ||||| |  | ||||| || ||  | ||||||||
Db       1872 GTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTGTCTGGCGCTCC 1813

Qy        883 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 942
              || || ||||| || |||||| | ||||||||  |  |  | ||||||||| || | ||
Db       1812 ATGGAACACCCGGGGAAGCTCGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 1753

Qy        943 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTT 1002
              ||||| || ||||| ||  |||| ||||||||||||| | || || ||  |      ||||
Db       1752 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTGGGTTC 1693

Qy       1003 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1062
              ||   | | || || ||      |||| | | | ||  ||||  |||| ||  |||| |||
Db       1692 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1633

Qy       1063 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1119
              ||| |  ||||| |   ||||  ||   |      ||  |   ||      |   | |
Db       1632 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1573

Qy       1120 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1179
                |   | ||| ||    | || ||  ||||| |     |||| |||||    ||| |
Db       1572 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1513

Qy       1180 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1239
                |    ||||||||        ||||| ||| | |||||  |  | || |||||  || ||
Db       1512 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCGG 1453

Qy       1240 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCA 1299
              ||   |||||||| ||||||||  |||||   ||||||||| ||||| || || || ||
Db       1452 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1393

Qy       1300 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1342
                | ||||||||||| |||||||| || ||||||||   ||||
Db       1392 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 1350
```

```
RESULT 40
US-09-908-153B-41
; Sequence 41, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beerli, Roger
;  APPLICANT:  Schopfer, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41
;   LENGTH: 2775
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-153B-41

  Query Match             18.5%;  Score 265.4;  DB 9;  Length 2775;
  Best Local Similarity   64.4%;  Pred. No. 2.1e-71;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    3;  Gaps    1;

Qy        703 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 762
             ||||||| || ||||| ||||||   |   ||||||| | || || | |||| ||||
Db        748 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 807

Qy        763 CACATGATCAGCTGGGCCAAGAACATTCCCGGCTTTGTGCAGCTCAGCCTGTTCGACCAA 822
             || ||||||| ||||||  |||   | || ||||||| ||| | | ||   || ||
Db        808 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATGATCAG 867

Qy        823 GTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCA 882
             || |  ||  | ||| |    ||| ||||| |  | ||||| || ||  | ||||||||
Db        868 GTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 927

Qy        883 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 942
             || || |||| || |||||| | ||||||||  |  |  | |||||||| || | ||
Db        928 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 987

Qy        943 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTT 1002
             ||||| || |||| ||  |||| ||||||||||| | || || ||  |    ||||
Db        988 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 1047

Qy       1003 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1062
             ||   | | || || ||     |||| | | | ||  ||||  |||| ||  |||| |||
Db       1048 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1107

Qy       1063 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1119
             ||| |  |||| |    |||  ||   |  |   ||  |    |  |  |  |  |||
Db       1108 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1167

Qy       1120 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1179
                |  | ||| ||   | || ||  ||||| |    |||| |||||    ||| |
Db       1168 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1227

Qy       1180 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1239
                |    ||||||||        ||||| ||| | |||||  |  | || |||||  || ||
Db       1228 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 1287

Qy       1240 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAAGTGTGGTCCCA 1299
             ||   |||||||| |||||||| |||||   |||||||| ||||| || || || ||
Db       1288 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1347

Qy       1300 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1342
                | ||||||||||| ||||||||| || ||||||||   ||||

Db       1348 CTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGGCTTC 1390


RESULT 41
US-09-908-153B-43/c
; Sequence 43, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
;  APPLICANT: Barbas, Carlos F.
;  APPLICANT:  Beerli, Roger
;  APPLICANT:  Schopfer, Ulrich
;  TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
;  TITLE OF INVENTION:  SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
;  TITLE OF INVENTION:  SWITCHES
;  FILE REFERENCE: TSRI 725.1
;  CURRENT APPLICATION NUMBER: US/09/908,153B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 09/619,063
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
;   LENGTH: 2775
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Synthesized
US-09-908-153B-43

  Query Match             18.5%;  Score 265.4;  DB 9;  Length 2775;
  Best Local Similarity   64.4%;  Pred. No. 2.1e-71;
  Matches  414;  Conservative    0;  Mismatches  226;  Indels    3;  Gaps    1;

Qy        703 CCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTA 762
             ||||||| || ||||| ||||||   |   ||||||| | || || | |||| ||||
Db       2028 CCCTTCAGTGAAGCCTCAATGATGGGCTTATTGACCAACCTAGCAGATAGGGAGCTGGTT 1969

Qy        763 CACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAA 822
             || ||||||| |||||| ||||   | || ||||||| ||| | | ||   || ||
Db       1968 CATATGATCAACTGGGCAAAGAGAGTGCCAGGCTTTGGGGACTTGAATCTCCATGATCAG 1909

Qy        823 GTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCA 882
             || |  ||  | ||| |    ||| ||||| |  | ||||| || ||  | ||||||||
Db       1908 GTCCACCTTCTCGAGTGTGCCTGGCTGGAGATTCTGATGATTGGTCTCGTCTGGCGCTCC 1849

Qy        883 ATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGG 942
             || || |||| || |||||| | ||||||||  |  |  | |||||||| || | ||
Db       1848 ATGGAACACCCGGGGAAGCTCCTGTTTGCTCCTAACTTGCTCCTGGACAGGAATCAAGGT 1789

Qy        943 AAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTT 1002
             ||||| || |||| ||  |||| ||||||||||| | || || ||  |    ||||
Db       1788 AAATGTGTGGAAGGCATGGTGGAGATCTTTGACATGTTGCTTGCTACGTCAAGTCGGTTC 1729

Qy       1003 CGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAAT 1062
             ||   | | || || ||     |||| | | | ||  ||||  |||| ||  |||| |||
Db       1728 CGCATGATGAACCTGCAGGGTGAAGAGTTTGTGTGCCTCAAATCCATCATTTTGCTTAAT 1669

Qy       1063 TCCAGTATGTACCC---TCTGGTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTG 1119
             ||| |  |||| |    |||  ||   |  |   ||  |    |  |  |  |  |||
Db       1668 TCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAAGAGAAGGACCACATC 1609

Qy       1120 GCTCACTTGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATC 1179
                |  | ||| ||   | || ||  ||||| |    |||| |||||    ||| |
Db       1608 CACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCTGGCCTG 1549

Qy       1180 TCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGG 1239
                |    ||||||||        ||||| ||| | |||||  |  | || |||||  || ||
Db       1548 ACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGG 1489

Qy       1240 CATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAAGTGTGGTCCCA 1299
             ||   |||||||| |||||||| |||||   |||||||| ||||| || || || ||
Db       1488 CACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTGCCC 1429
```

```
Qy        1300 GTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1342
               | |||||||||||| |||||||||| || ||||||||   ||||
Db        1428 CTCTATCACCTCCTCCTGGACATGTTCCATCCCCACCCCTTC 1386


RESULT 32
US-09-864-761-24607
; Sequence 24607, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24607
;   LENGTH: 306
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.8
;   OTHER INFORMATION, EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6
;   OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 9.00e-58
;   OTHER INFORMATION: EST_HUMAN HIT: BE510841.1, EVALUE 1.00e-46
;   OTHER INFORMATION: NT HIT: AB006590.1, EVALUE 0.00e+00

US-09-864-761-24607

  Query Match             14.2%;  Score 203;  DB 10;  Length 306;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-52;
  Matches  203;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 163

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 223

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAACATCGCTA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAACATCGCTA 283

Qy        181 CAACACACCTTACCTGTAAACAG 203
              |||||||||||||||||||||||
Db        284 GAACACACCTTACCTGTAAACAG 306


RESULT 33
US-09-864-761-7903
; Sequence 7903, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
```

```
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7903
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL     6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL   5.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL   4.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL   5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL   4.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL   4.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL     6
US-09-864-761-7903

  Query Match             14.2%;  Score 203;  DB 10;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 2e-52;
  Matches  203;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          252 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 311

Qy           61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          312 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 371

Qy          121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          372 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 431

Qy          181 GAACACACCTTACCTGTAAACAG 204
                |||||||||||||||||||||||
Db          432 GAACACACCTTACCTGTAAACAG 454


RESULT 44
US-09-864-761-6718
; Sequence 6718, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,412
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,466
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6718
;   LENGTH: 599
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161756.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL   1.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL   1.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL   1.1
US-09-864-761-6718

  Query Match             13.2%;  Score 189;  DB 10;  Length 599;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-48;
  Matches  189;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1246 AGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTAT 1305
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          366 AGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTAT 425

Qy         1306 GACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACG 1365
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          426 GACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACG 485

Qy         1366 GGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAG 1425
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          486 GGGTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAG 545

Qy         1426 TCTCAGTGA 1434
                |||||||||
Db          546 TCTCAGTGA 554


RESULT 45
US-09-864-761-23452
; Sequence 23452, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,412
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,466
;  PRIOR FILING DATE: 2000-08-04
```

```
;  PRIOR APPLICATION NUMBER: GB 2425[illegible]
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,459
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23452
;    LENGTH: 183
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: MAP TO AL161756.1
;    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;    OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
;    OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;    OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
;    OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;    OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;    OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
;    OTHER INFORMATION: NI HIT: X99101 1, EVALUE 6.00e-99
;    OTHER INFORMATION: EST_HUMAN HIT: AI174706 1, EVALUE 2.50e+00
;    OTHER INFORMATION: SWISSPROT HIT: Q92731, EVALUE 4.00e-30
US-09-864-761-23452

  Query Match             12.8%;  Score 183;  DB 10;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-46;
  Matches  183;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1248 TAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGA 1307
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATGA 60

Qy       1308 CCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGG 1367
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGGG 120

Qy       1368 GTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTC 1427
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCCGAGTGCAGCCCGGCAGAGGACAGTAAAAGCAAAGAGGGCTCCCAGAACCCACAGTC 180

Qy       1428 TCA 1430
             |||
Db        181 TCA 183


RESULT 36
US-09-962-436-268
; Sequence 268, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
;  APPLICANT: Soppet, Daniel
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-75
;  CURRENT APPLICATION NUMBER: US/09/962,436
;  CURRENT FILING DATE:  2001-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,082
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/234,924
;  PRIOR FILING DATE: 2000-09-25
;  NUMBER OF SEQ ID NOS: 568
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 268
;    LENGTH: 2218
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-962-436-268

  Query Match             9.3%;  Score 133.8;  DB 10;  Length 2218;
  Best Local Similarity   49.9%;  Pred. No. 1.2e-30;
  Matches  392;  Conservative    0;  Mismatches  387;  Indels    6;  Gaps    2;

Qy        280 CACTICTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
              | | |||||   |||||  | ||    || || || || |||||||| ||    || |||
Db        406 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 465

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db        466 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCG 525

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||  |||||   |||   ||     | |||||  ||||||||||||||  |
Db        526 GCCTCCAACGACTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 585

Qy        460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGACATGTGGGTAC 519
              |||||       ||||| ||| | |||   ||    ||    ||  |   ||
Db        586 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGT 645

Qy        520 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 579
              | |   | |   || | | | || | | | |   |     ||
Db        646 GGGCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTCCCGGCCCCTTC 705

Qy        580 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAG 639
              |||  ||  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        706 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCCCCAGTGAATGCACTG 765

Qy        640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 699
              |    ||    |  | ||| |||||  || | |||   ||    || | | |
Db        766 GTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGC 825

Qy        700 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 759
              |    |  |  | || ||  |  ||||    |  |  |||  ||| |
Db        826 CCTGATGGGCACCTCCCAGCCGTGGCTA---CCCTCTGTGACCTCTTTGACCGAGAGATT 882

Qy        760 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
              ||   || |||||||||||||||  || || |||||      |||   |||| |||
Db        883 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 942

Qy        820 CAAGTGGGGCTCTTGGAGAGCTGTTGGATGCAGGTGTTAATGATGGGGCTGATCTGGGGC 879
              ||  ||    |  || ||||   |||||||||||| |  || |||  ||     ||||
Db        943 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGGTGGGCCAGCGC 1002

Qy        880 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAG 939
              ||| |   |  | |   |||| ||| |||   ||  | || ||||  ||  |||
Db       1003 TCACTGCCACTGCAGGATGAGCTGCCTTGGCTGAGGACTTAGTCCTGGA---TGAAGAG 1059

Qy        940 GGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGG 999
              ||  | | | ||  ||  |  |||| |    |     || |||  |             ||
Db       1060 GGGGCAGGGCAGCTGGCCTGGGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGGG 1119
```

```
Qy        1000 TTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTC 1059
               | | |   |    ||  | |   | || ||| |     | |||||| ||  |||   |
Db        1120 CTGCAGGCCCTGCGGCTGGAGCGAGAGGAGTATGTTCTACTAAAGGCCTTGGCCCTTGCC 1179

Qy        1060 AATTC 1064
               |||||
Db        1180 AATTC 1184


RESULT 37
US-10-211-239-7
; Sequence 7, Application US/10211239
; Publication No. US20030104965A1
; GENERAL INFORMATION:
;  APPLICANT: Junq, Birqit
;  APPLICANT:  Kraut, No. US20030104965Albert
;  APPLICANT:  Mueller, Stefan
;  TITLE OF INVENTION: Method for Identifying Substances Which Positively
;  TITLE OF INVENTION:  Influence Inflammatory Conditions
;  FILE REFERENCE: 0652.2440001
;  CURRENT APPLICATION NUMBER: US/10/211,239
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/415,775
;  PRIOR FILING DATE: 2001-08-40
;  PRIOR APPLICATION NUMBER: EP 01119003.0
;  PRIOR FILING DATE: 2001-08-06
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 2402
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-211-239-7

  Query Match              9.0%;  Score 129;  DB 9;  Length 2402;
  Best Local Similarity   48.7%;  Pred. No. 4.8e-29;
  Matches  382;  Conservative    0;  Mismatches  400;  Indels    3;  Gaps    1;

Qy         280 CACTTCTGCGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
               | | |||||   |||||  | ||    || || || || ||||||| ||    || |||
Db         583 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 642

Qy         340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
               || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db         643 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCG 702

Qy         400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
               ||  | ||  ||||   |||   ||     | || |||  ||||||||||||||  |
Db         703 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 762

Qy         460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTAC 519
               |||||       ||||| ||| | |||   ||    ||    ||  |       ||
Db         763 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGT 822

Qy         520 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 579
               | |    |  |    || |  |  | || | | | |  |       |       ||
Db         823 GGGCGGCAGAAGTACAAGCGGCGGCGCGAGGTGGACCCACTGCCCTTCCCCGGCCCCTTC 882

Qy         580 AGAAGTGCCGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAG 639
                  |||   ||  ||   |  |   |    | | |   |  |||         |
Db         883 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCGCAAGACAGCACGCCCAGTGAATGCA 942

Qy         640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 699
               | |  |    |   ||  |||  |      |   |       ||      ||||
Db         943 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCA 1002

Qy         700 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 759
               | ||        |  | |     ||   ||||     | |    |||  ||| |
Db        1003 GGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGACAGATT 1062

Qy         760 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
               ||  ||  ||||||||||||||||  || || |||||      |||    ||||  |||
Db        1063 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCCGAC 1122
```

```
Qy         820 CAAGTGCGGCTCTTGGACAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGC 879
               ||  ||    |  || ||||   ||||||||||||| |  || ||||  ||     ||||
Db        1123 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGC 1182

Qy         880 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAG 939
               ||| |    |  |  |  ||||   ||| |||   ||  | || |||||    || |||
Db        1183 TCACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTAGTCCTGGA   TGAAGAG 1239

Qy         940 GGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGG 999
               |||  | | | ||  ||  |   |||| |    |      || |||  |          ||
Db        1240 GGGGCACGGGCAGCTGGCCTGGGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGCGG 1299

Qy        1000 TTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTC 1059
               | |  |    |    ||  | |   | || ||| |     | |||||| ||  |||   |
Db        1300 CTGCAGGCCCTGCGGCTGGAGCGAGAGGAGTATGTTCTACTAAAGGCCTTGGCCCTTGCC 1359

Qy        1060 AATTC 1064
               |||||
Db        1360 AATTC 1364


RESULT 38
US-10-037-270-454
; Sequence 454, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030104529Alel Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,417
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_Fi_qenes Version 1.0
; SEQ ID NO 454
;   LENGTH: 2539
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (97)..(1755)
US-10-037-270-454

  Query Match              9.0%;  Score 129;  DB 9;  Length 2539;
  Best Local Similarity   48.7%;  Pred. No. 4e-29;
  Matches  382;  Conservative    0;  Mismatches  400;  Indels    3;  Gaps    1;

Qy         280 CACTTCTGCGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
               | | |||||   |||||  | ||    || || || || ||||||| ||    || |||
Db         712 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 771

Qy         340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
```

```
Db        772 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCG 831

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCCCCGCAAGAGCTGCCAGCCTCCCCACTT 459
              ||  | ||  |||||   |||   ||     | ||||||  ||||||||||||||||  |
Db        832 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 891

Qy        460 CGGAAGTGTTACGAACTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTAC 519
              |||||       ||||| ||| | |||   ||     ||   |      ||
Db        892 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGCAGGAGTGCGCCTGGACCGCCTGCCGCCT 951

Qy        520 CGCCTTGTGCGGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 579
               | |    |  |    || |  |  | || | | |  |        |        ||
Db        952 GGGCGGCAGAACTACAAGCGGCGGCCGCAGGTGGACCCACTGCCCTTGCCGGGGCCGTTC 1011

Qy        580 AGAAGTGCCGGCCAGGCGGCGGAGTGCGGCAGCTGCTGCTGGACCGCCTGCAGCGGCAC 639
                 |||   ||  ||    |   | |  |  ||||    |  |||||||||||||||||
Db       1012 CCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCA 1071

Qy        640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTCATCAGCGGCCCAGT 699
              | | |    |   ||  |||  |    |    |  |                ||||
Db       1072 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCA 1131

Qy        700 GCGCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCTGACAAGGAGCTGG 759
              | ||        |   | |     ||   ||||     | |    |||   ||| |
Db       1132 GGCCGTGATGGGCAACATCAAGAACGAGCTCATCCACCTGTTGCACCTCCTTGACCGCAACAAT 1191

Qy        760 GTACACATGATCAGCTGGCCAAGAAGATTCCCGGCTTTGTGGACCTCACCCTGTTCGAC 819
              ||   ||  ||||||||||||||||  || ||  |||||      |||    ||||  |||
Db       1192 GTGGTCACCATCAGCTGGCCAAGAGCATCCCAGGCTTCTCATCCTTCTCCCTCTCTCAC 1251

Qy        820 CAAGTCCGGCTCTTCGAGAGCTGTTCGATGGAGGTCTTAATGATCCCCCTGATCCGGCCC 879
              ||  ||    |  || ||||   ||||||||||| |  || ||||  ||     ||||
Db       1252 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTCTGCTCCTGGGTCTCCCCAGCCC 1311

Qy        880 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAG 939
              ||| |    |  |  |  ||||   ||| |||   ||  | || ||||    || |||
Db       1312 TCACTGCCACTGCAGGATGAGCTGGCCTTCCGCTGAGGACTTAGTCCTGGA---TGAAGAG 1368

Qy        940 GGGAAATCCGTAGAAGGAATTCTCGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGT 999
              |||  | | |  ||  |   |||| |    |     || |||  |           ||
Db       1369 GGGGCACGGGCAGCTGGCCTGGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCCCCC 1428

Qy       1000 TTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGCCCATGATCCTGCTC 1059
               | |  |   |    ||  | |   | || ||| |     | ||||||  ||  ||   |
Db       1429 CTGCAGGCCCTGCCGCCTGGAGCGAGAGCAGTATGTTCTACTAAAGCCCTTGCCCTTGCC 1488

Qy       1060 AATTC 1064
              |||||
Db       1489 AATTC 1493


RESULT 39
US-10-008-739A-1
; Sequence 1, Application US/10008739A
; Patent No. US20020161194A1
; GENERAL INFORMATION:
;  APPLICANT: Pfizer Inc.
;  APPLICANT:  Castleberry, Tessa A.
;  APPLICANT:  Lu, Bihong
;  APPLICANT:  Owen, Thomas A.
;  APPLICANT:  Smock, Steven L.
;  TITLE OF INVENTION: The Canine Androgen Receptor
;  FILE REFERENCE: PC10893AGPR
;  CURRENT APPLICATION NUMBER: US/10/008,739A
;  CURRENT FILING DATE:  2002-04-15
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;  LENGTH: 4577
;  TYPE: DNA
;  ORGANISM: Canine
US-10-008-739A-1

  Query Match             6.8%;  Score 97.2;  DB 9;  Length 4577;
  Best Local Similarity   51.7%;  Pred. No. 3.5e-19;
  Matches  382;  Conservative    0;  Mismatches  298;  Indels   22;  Gaps    4;

Qy        293 TCTGCAGCCATTACCCATCGGCATATCACTATGGAGTCTGCTCGTCGTGAAGCATCTAAGG 352
              ||||| | ||| | || || || | ||||||||||       | |||| | | || || |
Db       1646 TCTGCGGTGATGAAGCTTCTGGCTGTCACTATGGAGCTCTCACTTGTGGAAGCTGCAAAG 1705

Qy        353 CCTTTTTTAAAAGAAGCATTCAAGGACATAATCATTATATTTGTCCAGCTACAAATCAGT 412
               ||| |||||||||  |  | ||||  |  |  | ||  | ||| |    | |||| | |
Db       1706 TCTTCTTTAAAAGAGCCGCTGAAGGGAAACAGAAGTACCTGTGTGCCAGCAGAAATGATT 1765

Qy        413 GTACAATCGATAAAAACCCCCGCAAGAGCTGCCAGGCCTGCCGCAGTTGGAAGTGTTACG 472
              |||| |||||||||  |||  | || |  || |   | || || |||| || || | |
Db       1766 GTACCATCGATAAATTCCGAAGGAAAAATTGTCCATCTTGTGCGCCTCCGGAAATGCTATG 1825

Qy        473 AAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGC--G 530
              |||  ||  |||     | | | |     |  | |||  |   ||| | |      |||  |
Db       1826 AAGCAGGGATGACTCTGGGAGCCCGGAAGCTAAAGAAACTGGGGAATCTGAAACTGCAAG 1885

Qy        531 GAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGG 590
               ||  |||| |   | |    || | || |       ||      |     | || |||
Db       1886 AGGAAGGAGAGGCTTCCAATGTCACCAGCCCCACTGAGGAGCCAACCCAGAAGCTGACGG 1945

Qy        591 CCAGGGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCT 650
              | | |    ||  ||   ||                | |||||      | ||
Db       1946 TGTCACACATTGAAGGCTATGAGTG----------TCAGCCCATCTTTCTGAATGTC 1992

Qy        651 CACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGGGCCCTTCAC 710
              |    |   |   || ||    |||||||  ||   |  ||    ||||  | |
Db       1993 CTTCAACCCATCCACCCAGCCGTGGTGTGTGCTGGACA-TCACAACAACCACCCCCACTC 2051

Qy        711 CGAGGCCTCCATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGCTGGTACACATGATT 770
              |   ||  || || |      ||| |   | ||||  || | | || | | ||||  || |
Db       2052 CTTCCAAGCCTCCGTCCTCAGCCTTAATGAATTGGGTCAAAGCCAGCTTGTACATGTGCT 2111

Qy        771 CAGCCGGGCCAAGAAGATTCCCGGCTTTGTGGACCTCACCCTGTTCGACCAAGTGCGGCT 830
              ||  |||||||||     | || |||||      | ||   | ||   |||||  ||    |
Db       2112 CAAGTGGCCCAAGGCCTTGCCGGGCTTCCGCAACCTGCACGTGGATGACCAGATGGCAGT 2171

Qy        831 CTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCG------CTCAAT 884
              | |  ||  ||  |||||| | |  |  || | |   ||  |||||      | | |
Db       2172 CATTCAGTACTGCTGGATGCGGCTCATGGTGTTTGCCATGGGCTGGCGATCCTTCACCAA 2231

Qy        885 TGACCACCCCGGCCAAGCTCATCTTTGCTCCAGATCTTGTTCT 926
              || | ||  || | | ||||  ||| || || || || ||| |
Db       2232 TGTCAACTCCAGGATGCTCTACTTCGCCCCTGACCTGGTTTT 2273


RESULT 40
US-09-997-267-1
; Sequence 1, Application US/09997267
; Patent No. US20020165381A1
; GENERAL INFORMATION:
;  APPLICANT: AHRENS-FATH, ISABELLE
;  APPLICANT:  HAENDLER, BERNARD
;  TITLE OF INVENTION: HUMAN ANDROGEN RECEPTOR VARIANTS
;  FILE REFERENCE: SCH-1793
;  CURRENT APPLICATION NUMBER: US/09/997,267
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: 60/255,078
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: Patent'n Ver. 2.1
; SEQ ID NO 1
;  LENGTH: 1329
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-997-267-1

  Query Match             6.6%;  Score 95.2;  DB 9;  Length 1329;
  Best Local Similarity   51.2%;  Pred. No. 8.4e-19;
```

```
  Matches  342;  Conservative    0;  Mismatches  298;  Indels   18;  Gaps    4;

Qy        285 CTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGG 344
              ||||    ||||  | ||| | || || || | ||||||||||        | |||| | |
Db        244 CTGCCTGATCTGTGGAGATGAAGCTTCTGGGTGTCACTATGGAGCTCTCACATGTGGAAG 302

Qy        345 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 404
              || |||| ||| || |||||  |  | ||||  |  |  | ||  | ||  |    |
Db        303 CTGCAAGGTCTTCTTCAAAAGAGCCGCTGAAGGGAAACAGAAGTACCTGTGCGCCAGCAG 362

Qy        405 AAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 464
              |||| | || || || ||||||  |||  | || |  || |   | || || |||||||
Db        363 AAATGATTGCACTATTGATAAATTCCGAAGGAAAAATTGTCCATCTTGTCGTCTTCGGAA 422

Qy        465 GTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGTAGAGAGAGATGTGGGTACCGCCT 524
              |||||  |||| || |||     | | | |     |   ||| |  ||| | | |
Db        423 ATGTTATGAAGCAGGGATGACTCTGGGAGCCCGGAAGCTGAAGAAACTTGGTAATCTGAA 482

Qy        525 TGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAG 584
              |   | ||   || || | |  | |||| |  ||      | |    | |||   |
Db        483 ACTACAG GAGGAAGGAGAGGCTTCCAGCACCACCAGCCCCACTGAGGAGACAACCCAGA 541

Qy        585 TGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCT 644
              |  | |  |  | | |      ||       ||          | |||||
Db        542 AGCTGACAGTGTCACACATTGAAGGCTATGAATG    ---- -TCAGCCCATCTTTCTG 591

Qy        645 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCC 704
              | ||  |     |    |    ||  |     |||||||  || |  |  ||    |||
Db        592 AATGTCCTGGAAGCCATTGAGCCAGGTGTAGTGTGTGCTGGACA CGACAACAACCAGCC 650

Qy        705 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACA 764
              |  | ||   ||  || || |     ||| |   |  |||  || |   || | ||||
Db        651 CGACTCCTTTGCAGCCTTGCTCTCTAGCCTCAATGAACTGGGAGACAGACAGCTTGTACA 710

Qy        765 CATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGT 824
              | || |||  |||||||||    | || ||||     |  |   | ||  ||||||  |
Db        711 CGTGGTCAAGTGGGCCAAGGCCTTGCCTGGCTTCCGCAACTTACACGTGGACGACCAGAT 770

Qy        825 GCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCG        878
              |    || |  ||  ||  |||||| | |  |  || | |   ||   |||||
Db        771 GGCTGTCATTCAGTACTCCTGGATGGGGCTCATGGTGTTTGCCATGGGCTGGCGATCCTT 830

Qy        879 CTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCT 926
              | | | || | || || | | |||  ||| || || |||| ||| |
Db        831 CACCAATGTCAACTCCAGGATGCTCTACTTCGCCCCTGATCTGGTTTT 878


Search completed: June 16, 2003, 13:06:24
Job time : 220.364 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 23:28:27 ; Search time 2182.44 Seconds
(without alignments)
10641.499 Million cell updates/sec

Title: US-08-826-361A-1
Perfect score: 1434
Sequence: 1 ATGAATTACAGCATTCCCAG . . . AGAACCCACAGTCTCAGTGA 1434

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8094743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	262.4	18.3	727	12	BG722442	BG722442 602692671
c 2	208	14.5	627	14	BM942391	BM942391 UI-M-BH3-
c 3	162.4	11.3	528	14	BM933826	BM933826 UI-M-BH3-
4	162.2	11.3	580	12	BG738138	BG738138 tp04105.y
5	157	10.9	642	12	BF058527	BF058527 7k31b09.x
6	155.4	10.8	641	10	AW976836	AW976836 ES1488745
7	153.8	10.7	648	12	BF660250	BF660250 maa95d12.
8	149.6	10.4	642	12	BF733032	BF733032 nae18e05.
9	148.2	10.3	782	10	AW134052	AW134052 ES14E07.y
10	131.4	9.2	594	12	BG084311	BG084311 H3049A08-
11	129.8	9.1	1142	14	BM810012	BM810012 AGENCOURT
12	126.4	8.8	137	17	AF455798	AF455798 AF455798
c 13	121	8.4	558	12	BG891892	BG891892 fp46e11.x
14	117	8.2	692	13	BJ068905	BJ068905 BJ068905
15	115.6	8.1	532	17	AZ596395	AZ596395 IM0409D10
16	115.2	8.0	611	17	AQ719425	AQ719425 HS_5512_B
17	110.6	7.7	1443	11	BC008218	BC008218 Homo sapi
18	108.8	7.6	618	13	BJ069238	BJ069238 BJ069238
19	107.8	7.5	703	12	BG744031	BG744031 602722811
c 20	107.4	7.5	575	12	BF510841	BF510841 UI-H-BI4-
21	107.4	7.5	2199	11	AK019222	AK019222 Mus muscu
22	106.2	7.4	676	12	BG747925	BG747925 602705778
23	102	7.1	844	13	BI100868	BI100868 602886184
24	101.4	7.1	937	13	BM562894	BM562894 AGENCOURT
25	100.4	7.0	558	13	BJ058831	BJ058831 BJ058831
26	99.8	7.0	908	10	BF546976	BF546976 601071950
27	99.6	6.9	800	9	AU154990	AU154990 AU154990
28	99.6	6.9	1004	9	AL527012	AL527012 AL527012
29	98.6	6.9	775	12	BG480960	BG480960 602529118
30	98.4	6.9	979	12	BG033384	BG033384 602301556
31	95	6.6	818	13	BI738331	BI738331 603361587
32	93	6.5	402	12	BF717864	BF717864 id52h03.y
33	91.8	6.4	669	14	BM957249	BM957249 fy75h10.y
34	91.2	6.4	434	10	BB711891	BB711891 BB711891
35	91.2	6.4	521	12	BG687711	BG687711 334912 BA
36	91.2	6.4	698	10	BB196002	BB196002 BB196002
37	91.2	6.4	1038	11	BC016462	BC016462 Mus muscu
38	91	6.3	599	12	BG010872	BG010872 RC5-GN019
39	90.6	6.3	592	12	BG746447	BG746447 602703728
c 40	90.2	6.3	413	10	AW113877	AW113877 MC809 mou
41	90.2	6.3	1010	9	AL547737	AL547737 AL547737
42	89.6	6.2	475	10	BB196485	BB196485 BB196485
43	89.6	6.2	425	10	BB710863	BB710863 BB710863
44	89.2	6.2	1120	14	BM908889	BM908889 AGENCOURT
45	88.8	6.2	729	13	BG914551	BG914551 602813303
46	88.2	6.2	462	9	AA411285	AA411285 zv26g03.r
c 47	88.2	6.2	477	9	AI341113	AI341113 qx84c02.x
c 48	88.2	6.2	479	9	AI653325	AI653325 wb24c06.x
49	88.2	6.2	549	9	AL047238	AL047238 DKFZp586N
50	88.2	6.2	558	14	BM786552	BM786552 K-EST0065
c 51	88.2	6.2	644	12	BF223014	BF223014 7q27e02.x
52	88.2	6.2	754	12	BG682389	BG682389 602640127
53	88.2	6.2	889	9	AL561541	AL561541 AL561541
54	88.2	6.2	891	9	AL525589	AL525589 AL525589
55	88.2	6.2	909	14	BQ898227	BQ898227 AGENCOURT
56	88.2	6.2	926	9	AL513888	AL513888 AL513888
57	88.2	6.2	960	14	BQ879017	BQ879017 AGENCOURT
58	88.2	6.2	967	9	AL555839	AL555839 AL555839
59	88.2	6.2	972	14	BQ882678	BQ882678 AGENCOURT
60	88.2	6.2	1019	13	BM557938	BM557938 AGENCOURT
61	88.2	6.2	1090	12	BG176463	BG176463 602338137
62	86.8	6.1	674	14	BM791802	BM791802 K-EST0071
63	86.8	6.1	911	14	BQ948248	BQ948248 AGENCOURT
64	86.6	6.0	1176	12	BF162189	BF162189 601770663
65	86.4	6.0	385	10	AV647429	AV647429 AV647429
66	86.4	6.0	399	10	AV647483	AV647483 AV647483
67	86.4	6.0	399	10	AV647486	AV647486 AV647486
c 68	86.4	6.0	446	9	AI636264	AI636264 tz93c10.x
c 69	86.4	6.0	465	9	AI393892	AI393892 tq05c05.x
c 70	86.4	6.0	519	9	AU143800	AU143800 AU143800
71	86.4	6.0	618	12	BG723061	BG723061 602690613
72	86.2	6.0	863	14	BQ949290	BQ949290 AGENCOURT
73	86	6.0	416	13	BI133725	BI133725 UI-M-BH3-
74	86	6.0	794	12	BF138044	BF138044 601784742
c 75	85.8	6.0	271	13	BI300602	BI300602 UI-R-CV2-
76	85.8	6.0	450	17	AZ032574	AZ032574 RPCI-23-3
77	85.8	6.0	992	9	AL547687	AL547687 AL547687
78	85.8	6.0	1071	14	BM906677	BM906677 AGENCOURT
79	85.8	6.0	3141	11	BC030234	BC030234 Homo sapi

c	80	85.6	6.0	279	12	BG474797	BG474797 UI-R-CV1
c	81	85.6	6.0	279	12	BI298646	BI298646 UI-R-CV2
	82	85.6	6.0	569	10	AV968271	AV968271 AV968271
c	83	85.4	6.0	758	14	BQ446111	BQ446111 UI-H-EU1-
	84	85.2	5.9	645	10	BE533846	BE533846 601237423
	85	85.2	5.9	947	9	AL527466	AL527466 AL527466
	86	85	5.9	910	9	AL556302	AL556302 AL556302
	87	85	5.9	2168	12	AK004789	AK004789 Mus musculu
c	88	84.8	5.9	445	12	BE551955	BE551955 UI-R-C2p-
c	89	84.8	5.9	484	9	AA851819	AA851819 EST194087
c	90	84.8	5.9	484	9	AA893618	AA893618 EST197421
c	91	84.8	5.9	490	14	WI3715	WI3715 ma83b04.r1
	92	84.8	5.9	428	9	AI323943	AI323943 ma83b01.x
c	93	84.8	5.9	451	9	AI327473	AI327473 ma83b04.y
	94	84.8	5.9	500	10	AW490389	AW490389 UI-M-BH3-
	95	84.8	5.9	574	9	AI171721	AI171721 EST217703
	96	84.8	5.9	588	10	AW491078	AW491078 UI-M-BH3
	97	84.8	5.9	655	10	AV647680	AV647680 AV647680
	98	84.8	5.9	963	14	BQ936063	BQ936063 AGENCOURT
	99	84.2	5.9	466	9	AA744644	AA744644 ny79f02.s
c	100	84.2	5.9	610	14	BM941929	BM941929 UI-M-CC0p
	101	84.2	5.9	621	12	BF143483	BF143483 601790832
c	102	84.2	5.9	972	17	CNS04TUQ	AL390971 Tetraodon
c	103	84	5.9	276	13	BI300572	BI300572 UI-R-CV2-
c	104	84	5.9	955	17	CNS04QE6	AL392487 Tetraodon
	105	84.8	5.8	783	9	AL561358	AL561358 AL561358
	106	84.8	5.8	1103	14	BM908460	BM908460 AGENCOURT
	107	84.8	5.8	1114	12	BG031969	BG031969 602300664
c	108	84.6	5.8	714	17	CNS01VNY	AL169351 Tetraodon
c	109	84.2	5.8	259	13	BI299473	BI299473 UI-R-CV2
	110	84.2	5.8	750	13	BI759107	BI759107 603042512
	111	84.2	5.8	818	13	BM020032	BM020032 603648565
	112	82.8	5.8	544	9	AA517920	AA517920 vh95c10.r
	113	82.6	5.8	663	13	BI832037	BI832037 603075630
	114	82.6	5.8	682	13	BI827058	BI827058 603075729
	115	82.2	5.7	417	10	BB276737	BB276737 BB276737
	116	82.2	5.7	431	10	BE665969	BE665969 155233 MA
	117	82.2	5.7	450	12	BF937961	BF937961 fm70q06.y
	118	82.2	5.7	524	12	BG482331	BG482331 602526670
	119	82.2	5.7	580	12	BG425312	BG425312 602423886
	120	82.2	5.7	629	14	BM739741	BM739741 K EST0215

ALIGNMENTS

```
RESULT 1
BG722442
LOCUS       BG722442                727 bp    mRNA    linear   EST 08 MAY 2001
DEFINITION  602693671F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4825937 5',
            mRNA sequence.
ACCESSION   BG722442
VERSION     BG722442.1  GI:14091629
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10739  row: n  column: 18
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1..727
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4825937"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI XhoI (qtcgag
                     ); Oligo-dT primed using primer 5' TTTTTTTTTTTTTTTVN-3',
                     size selected for  average insert size 2.2 kb and
                     normalized to EOT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      182 a    194 c    175 g    176 t
ORIGIN

  Query Match           18.3%;  Score 262.4;  DB 12;  Length 727;
  Best Local Similarity   99.3%;  Pred. No. 5e-63;
  Matches  274;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 510

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 570

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 630

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAA-GGAAGGTTAGTGGGAACCGTTC 239
              |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||
Db        631 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAACGGACGGTTAGTGGGAACCGTTG 690

Qy        240 CGCCAGCCCTGTTACTGGTCCAGGTTCAAAGACGGA 275
              ||||||||||||||||||||||||||||||||||||
Db        691 CGCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGA 726


RESULT 2
BM942391/c
LOCUS       BM942391                627 bp    mRNA    linear   EST 13 MAR 2002
DEFINITION  UI-M-BH3-bsc-b-09-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-bsc-b-09-0-UI 5', mRNA sequence.
ACCESSION   BM942391
VERSION     BM942391.1  GI:19402254
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, B
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9644, Bethesda, MD
            20892-9644, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
```

```
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-bsc-b-09-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately derived from a mixture of individually tagged
                     normalized libraries from ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland, striatum,
                     hipoccampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows: PCR-amplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S4 library. This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
BASE COUNT       147 a    151 c    184 g    144 t      1 others
ORIGIN

  Query Match             14.5%;  Score 208;  DB 14;  Length 627;
  Best Local Similarity   63.3%;  Pred. No. 1.2e-47;
  Matches  335;  Conservative    0;  Mismatches  191;  Indels    3;  Gaps    1;

Qy        817 GACCAAGTGGGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGCCCTGATTGCC 876
              || || || |  ||  | |||  |    ||| ||||| |  | ||||| || ||  | |||
Db        624 GATCAGGTCCACCTTCTGGAGTGTGCCTGCCTCCAGATTCTGATGATTGGTCTCGTCTGC 565

Qy        877 CGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCAT 936
              ||||| || || ||||| |  |||||| | | ||||||  |  |  | |||||||||| ||
Db        564 CGCTCCATGGAACACCCGGAGAAGCTCCTGTNTGCTCCTAACTTGCTCCTGGACAGGAAT 505

Qy        937 GAGGGAAATGGGTAGAAGCAATTCTGGAAATCTTTGACATGCTCCTGCCAACTACTTCA 996
              | || ||||| || ||||| ||  |||| ||||||||||||| | ||||| ||  |
Db        504 CAAGGTAAATGTGTGGAAGGCATGGTGGACATCTTTGACATGTTGCTGGCTACGTCAAGT 445

Qy        997 AGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATGCTG 1056
              |||| ||  | || || ||     |||| | | | ||  |||| |||| ||  ||
Db        444 CGGTTCCGCATGATGAACCTGCAGGGAGAAGAGTTTGTGTGCCTCAAATCCATCATTTTG 385

Qy       1057 CTCAATTCCAGTATGTACCC---TCTCCTCACAGCGACCCAGCATGCTGACAGCACCCCC 1113
              || |||||| |  |||| |    |||  ||   |     ||  |   ||    |
Db        384 CTTAATTCCGGAGTGTACACGTTTCTGTCCAGCACCTTGAAGTCTCTGGAACACAACGAC 325

Qy       1114 AAGCTGGCTCACTTGCTGAACGCCGTGACCCATGCTTTGGTTTGGGTCATTGCCAAGACC 1173
              |  |    |  | ||| ||    | || ||  |||| |     |||| ||||
Db        324 CACATCCACCGTGTCCTGGACAAGATCACAGACACTTTGATCCACCTGATGGCCAAAGCT 265

Qy       1174 GGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCAC 1233
              ||| |  |   |||||||        ||||| ||| | |||||  |  | || |||||
Db        264 GGCCTGACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCAT 205

Qy       1234 GTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATGAAGTGCAAAAATGTG 1293
              || ||||   |||||||||| |||||||| |||||   |||||||||| ||||| || ||
Db        204 ATCCGGCACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTT 145

Qy       1294 GTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCCACGTGCTTC 1342
              || ||  | |||||||||| |||||||| || ||||||||   ||||
Db        144 GTGCCCCTCTATGACCTGCTCCTGGAGATGTTGGATGCCCACCGCCTTC 96


RESULT 3
BM933826/c
LOCUS       BM933826                 528 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  UI-M-BH3-bsc-b-09-0-UI.r2 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-bsc-b-09-0-UI 5', mRNA sequence.
ACCESSION   BM933826
VERSION     BM933826.1  GI:19992978
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-bsc-b-09-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately derived from a mixture of individually tagged
                     normalized libraries from ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland, striatum,
                     hipoccampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows: PCR-amplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S4 library. This procedure has been previously
```

described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)"
BASE COUNT 119 a 123 c 156 g 128 t 2 others
ORIGIN

Query Match 11.8%; Score 162.4; DB 14; Length 528;
Best Local Similarity 62.7%; Pred. No. 9.3e-45;
Matches 269; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

```
Qy        917 ATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATTCTGGAAATCTTTGACA 976
              |  |  | ||||||||| || | || ||||| || ||||| ||  |||| ||||||||||
Db        525 ACNTGCTCCTGGACAGGAATCAAGGTAAATGTGTCGAAGGCATGGTGGAGATCTTTGACA 466

Qy        977 TGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCT 1036
              || | ||||| ||  |     |||| ||   | | || || ||     |||| | | | |
Db        465 TGTTGCTGGCTACGTCAAGTCGGTTCCGCATCATCAACCTGCAGGGAGAAGAGTTTGTCT 406

Qy       1037 GTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCC---TCTGGTCACAGCGACCC 1093
              |  |||| |||| ||  |||| ||||||  |  ||||| |   |||| ||    |
Db        405 GCCTCAAATCCATCATTTTGCTTAATTCCGGAGTCTACACGTTTCTGTCCAGCACCTTGA 346

Qy       1094 AGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTGG 1153
              ||  |   ||     |    |  |     |  ||| ||    | || ||  |||||
Db        345 AGTCTCTGGAAGAGAAGGACCACATCCACCGTGTCCTGGACAAGATCACAGACACTTTGA 286

Qy       1154 TTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACC 1213
              |     |||| |||||    ||| |  |    |||||||| |       ||||| ||| | |
Db        285 TCCACCTGATGGCCAAAGCTGGCCTGACTCTGCAGCAGCAGCATCGCCGCCTAGCTCAGC 226

Qy       1214 TCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCA 1273
              |||| |  | || ||||  || ||||  |||||||| |||||||| ||||   ||
Db        225 TCCTTCTCATTCTTTCCCATATCCGGCACATGAGTAACAAAGGCATGGAGCATCTCTACA 166

Qy       1274 ACATGAAGTGCAAAAATGTGGTCCCAGTGTATGACCTGCTGCTGGAGATGCTGAATGCCC 1333
              ||||||| |||| || || || ||  | |||||||||| |||||||| || ||||||
Db        165 ACATGAAATGCAAGAACGTTGTGCCCCTCTATGACCTGCTCCTGGAGATGTTGGATGCCC 106

Qy       1334 ACGTGCTTC 1342
              ||   ||||
Db        105 ACCGCCTTC 97
```

RESULT 4
BG738138
LOCUS BG738138 580 bp mRNA linear EST 15-MAY-2001
DEFINITION tp04105.y1 zebrafish gridded kidney Danio rerio cDNA clone 4728705 5' similar to TR:O14012 O14012 ESTROGEN RECEPTOR. ;, mRNA sequence.
ACCESSION BG738138
VERSION BG738138.1 GI:14087827
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 580)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Other_ESTs: tp04105.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 477.
FEATURES Location/Qualifiers
source 1..580
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="4728705"
/clone_lib="zebrafish gridded kidney"
/sex="mixed"
/tissue_type "kidney pooled from 300 wild type adults"
/lab_host="XLOLR"
/note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; Oligo dT cDNA library constructed from mRNA pooled from pooled kidney tissue from 300 adult zebrafish."
BASE COUNT 133 a 161 c 156 g 130 t
ORIGIN

Query Match 11.8%; Score 162.2; DB 12; Length 580;
Best Local Similarity 76.0%; Pred. No. 1.1e-44;
Matches 200; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        268 AAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGA 327
              ||  | ||  |||||||||||||| || |||   || |||||| || || ||||||||||
Db        257 AAAGGAGACACTCACTTCTGCGCCGTGTGCCATGACTACGCTTCTGGGTATCACTATGGT 316

Qy        328 GTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGAT 387
              |||||||| |||||||| || |||| || || ||  | |||||||||| || |||||
Db        317 GTCTGGTCATGTGAAGGGTGCAAGGCTTTCTTCAAGCGTAGCATTCAAGGTCACAATGAC 376

Qy        388 TATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAG 447
              |||||||||||||| || || |||| || || || || | | | ||||||||||| |||
Db        377 TATATTTGTCCAGCCACCAACCAGTGCACTATTGACAAGAGCAGACGCAAGAGCTGTCAG 436

Qy        448 GCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAG 507
              ||||| |||| || |||| || |||||||| ||| ||||||||||     |||| |||
Db        437 GCCTGTCGACTCCGCAAGTGCTATGAAGTGGGCATGATGAAGTGTGGTGTTAGGAGGGAG 496

Qy        508 AGATGTGGGTACCGCCTTGTGCG 530
              | ||  | || ||   ||  ||
Db        497 CGCTGCAGTTATCGAGGTGCTCG 519
```

RESULT 5
BF058527
LOCUS BF058527 642 bp mRNA linear EST 16-OCT-2000
DEFINITION 7k81h08.x1 NCI_CGAP_ov18 Homo sapiens cDNA clone IMAGE:3477350 3' similar to SW:ESR1_HUMAN P03372 ESTROGEN RECEPTOR ;, mRNA sequence.
ACCESSION BF058527
VERSION BF058527.1 GI:10812423
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

```
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 492.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3477350"
                     /clone_lib="NCI_CGAP_Ov18"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGCCGCCGCACATTTTTTTTTTTTTTTTT 3']:
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      158 a    182 c    196 g    106 t
ORIGIN

  Query Match             10.9%;  Score 157;  DB 12;  Length 642;
  Best Local Similarity   77.6%;  Pred. No. 3.5e-33;
  Matches  190;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || ||||||||||
Db        346 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 405

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        406 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 465

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || |||||||||| || | ||||||||||||||
Db        466 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGCAGGAAGAGCTGCCAGGCC 525

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||| |  || | ||| ||
Db        526 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 585

Qy        511 TGTGG 515
              | ||
Db        586 AGAGG 590


RESULT 6
AW976636
LOCUS       AW976636                 641 bp    mRNA    linear   EST 02-JUN-2000
DEFINITION  EST388745 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW976636
VERSION     AW976636.1  GI:8167867
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 4528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 364
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGN"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      159 a    175 c    195 g    112 t
ORIGIN

  Query Match             10.8%;  Score 155.4;  DB 10;  Length 641;
  Best Local Similarity   77.1%;  Pred. No. 9.9e-33;
  Matches  189;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || ||||  || || || || || || || ||||||||||
Db        334 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 393

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| ||||| || || |||||||| || || ||||| ||||||||||||||| || |||
Db        394 TGGTCCTGTGAGGGCTGCAAGGCCTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 453

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || ||||| || || |||||||| || | ||||||||||||||
Db        454 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGGAGGAAGAGCTGCCAGGCC 513

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              ||||| || || || || ||||||||||||||| |||| |||| |  || | ||| ||
Db        514 TGCCGGCTCCGCAAATGCTACGAAGTGGGAATGATGAAAGGTGGGATACGAAAAGACCGA 573

Qy        511 TGTGG 515
              ||
Db        574 ATAGG 578


RESULT 7
BF660250
LOCUS       BF660250                 648 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  maa95q12.x1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3824830 3' similar to SW:ESR1_MOUSE P19785 ESTROGEN RECEPTOR
            ;, mRNA sequence.
ACCESSION   BF660250
VERSION     BF660250.1  GI:11925456
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MCI.1461182
            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3824830"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="thymus"
                     /dev_stage="4 weeks"
```

```
                     /lab_host "DH10B"
                     /note="Vector: pT7T4D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      158 a    184 c    183 g    123 t
ORIGIN

  Query Match           10.7%;  Score 153.8;  DB 12;  Length 648;
  Best Local Similarity   76.7%;  Pred. No. 2.8e-32;
  Matches  188;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || |||| || || || || || || || ||||| |||
Db        283 AAGGAGACTCGCTACTGTGCCGTGTGCAATGACTATGCCTCTGGCTACCATTATGGGGTC 342

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| || ||||| || ||||| || ||||| |||||||||||||||| ||||| ||
Db        343 TGGTCCTGCGAAGGCTGCAAGGCTTTCTTTAAGAGAAGCATTCAAGGACACAATGACTAC 402

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||||||||| || || || || || || |||||| | ||||| |||||||||
Db        403 ATGTGTCCAGCTACAAACCAATGCACCATTGACAAGAACCGGAGGAAGAGTTGCCAGGCC 462

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              || || || || ||||||||||||||||| ||| |||| | |||  |||| ||| |
Db        463 TGTCCGCTGCGCAAGTGTTACGAAGTGGGCATGATGAAAGGCGGCATACGGAAAGACCGC 522

Qy        511 TGTGG 515
                | ||
Db        523 CGAGG 527


RESULT 8
BF734032
LOCUS       BF734032                642 bp    mRNA    linear   EST 30-MAR-2001
DEFINITION  nae18e05.x1 NCI_CGAP_ov18 Homo sapiens cDNA clone IMAGE:3435729 3'
            similar to SW:ESR1_HUMAN P03372 ESTROGEN RECEPTOR ;, mRNA sequence.
ACCESSION   BF734032
VERSION     BF734032.1  GI:12058107
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer:  -40UP from Gibco
            High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:3435729"
                     /clone_lib "NCI_CGAP_ov18"
                     /tissue_type "fibrotheoma"
                     /lab_host "DH10B (phage resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCGACATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      156 a    184 c    193 g    109 t
ORIGIN

  Query Match           10.4%;  Score 149.6;  DB 12;  Length 642;
  Best Local Similarity   77.1%;  Pred. No. 4.4e-31;
  Matches  182;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        271 AGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTC 330
              | |||  ||| || ||| || || |||| || || || || || || || ||||||||
Db        346 AAGGAGACTCGCTACTGTGCAGTGTGCAATGACTATGCTTCAGGCTACCATTATGGAGTC 405

Qy        331 TGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTAT 390
              ||||| |||| || || |||||||| || || |||| ||||||||||||| || |||
Db        406 TGGTCCTGTGAGGGCTGCAAGGCTTTCTTCAAGAGAAGTATTCAAGGACATAACGACTAT 465

Qy        391 ATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCC 450
              || |||||||| || || |||| || || |||||||| || | ||||||||||| |||
Db        466 ATGTGTCCAGCCACCAACCAGTGCACCATTGATAAAAACAGCAGGAAGAGCTGCCAAGCC 525

Qy        451 TGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGA 506
              ||||| || || || || |||||| |||||||| |||| |||| || |||
Db        526 TGCCGGCTCCGCAAATGCTACGAAATGGGAATGATGAAAGGTGGGATACGATAAGA 581


RESULT 9
AW134052
LOCUS       AW134052                782 bp    mRNA    linear   EST 27-OCT-1999
DEFINITION  fi14h07.y1 Suqano Kawakami zebrafish DRA Danio rerio cDNA clone
            2601181 5' similar to qb:M12674 ESTROGEN RECEPTOR (HUMAN);, mRNA
            sequence.
ACCESSION   AW134052
VERSION     AW134052.1  GI:6135659
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopteryqii; Neopteryqii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 782)
  AUTHORS   Suqano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theisinq,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: fi14h07.x1
            Contact: S.L. Johnson
            Washinqton University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Suqano and Dr. Koichi Kawakami DNA
            Sequencing by: Washinqton University Genome Sequencinq Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1. .782
                     /organism "Danio rerio"
                     /strain "AB"
                     /db_xref "taxon:7955"
```

```
                     /clone "2601181"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGC);
                     Site_2: DraIII (CACCATGC); 1st strand cDNA was primed
                     with an oligo(dT) primer [AIGTGGCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor
                     [TGTTGGCCTACTGC], digested and cloned into plasmid between
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACGTGCAGCTCGAGCACA. "
BASE COUNT      195 a    230 c    176 g    179 t      2 others
ORIGIN

  Query Match             10.0%;  Score 143.2;  DB 10;  Length 782;
  Best Local Similarity   61.0%;  Pred. No. 3.2e-29;
  Matches  267;  Conservative    0;  Mismatches  165;  Indels    6;  Gaps    2;

Qy         18 CAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAATGTGTT 77
              |||||      ||   | |||  |   | |||  || ||| |   |||||| | | | ||
Db        351 CAGCACATCGCCITTIATCGGACTGCICGTCTGTGCGCCAGTCGCTAAGCCCGACTTTATT 410

Qy         78 GTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACATCTGTA 137
              |||||| | |   | ||  |||| |  |     | || |  |||     || ||   |
Db        411 TTGGCCACCTCACAGCCATGTTCCCCACTCACATTGCAACAACAGAGTCGACTTCAACA 470

Qy        138 TGCGCAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTACCTGT 197
                |  ||   |   |  | |||   |||   | | | |  || ||      |   |
Db        471 GAACCATGCTACTAGTGGGACTTGGACAGAACATACACCACATCATCATGTTGAAGAGGA 530

Qy        198 AAACAGAGACACACTGAAAAGCAAGGTTAGTGCGAACCCITCCGCCAGCCCTGTTACTGC 257
              ||||||  |  |||||   | |  || ||| |  || |  | | | |  |||  |||
Db        531 AAACAGCAAACCACTGGTGAAGCCGG-TAGCAGATACAGAGGACACGTCTGTGTCTCTGA 589

Qy        258 TCCAGGTTCAAAGAGGGATGCTCACTTCTGCCGCTGTCTGCAGCGATTACCCATCCGGATA 317
                |   ||    ||    |||| ||| || |||||||| |||||||  || |||||
Db        590 -----GAGGTAAAGCTGACATCCACTACTGTGCCGTCTGCAGTGATTACGCCTCTGGATA 644

Qy        318 TCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAACAAGCATTCAAGC 377
              ||| |||||||| ||||||||||||||| || || |||||||| || || || || ||||
Db        645 TCATTATGGAGTCTGGTCGTGTGAAGGCTGCAAAGCCTTTTTCAAGAGGAGTATACAAGG 704

Qy        378 ACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAA 437
              ||| |||| || || |||| || |||| || || || ||||  | |||||||||
Db        705 ACACAATGACTACATCTGTCCIGCCACAAACCAATGCACTATCGANCAGAACCGGCGCAN 764

Qy        438 GAGCTGCCAGGCCTGCCG 455
                |||||||||||||||||
Db        765 AAGCTGCCAGGCCTGCCG 782


RESULT 10
BG084311
LOCUS       BG084311                 594 bp    mRNA    linear   EST 26-JAN-2001
DEFINITION  H3099A08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3099A08 5', mRNA sequence.
ACCESSION   BG084311
VERSION     BG084311.1  GI:12566875
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Karqul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3099A08-3
            Contact: George J. Karqul
            Laboratory of Genetics
            National Institute on Aqinq/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit NIA Mouse cDNA Project Web site at http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3099  row: A  column: 08
            Seq primer: -21M13 Reverse
            High quality sequence stop: 594
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..594
                     /organism="Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref="nia EST:H3099A08-5"
                     /db_xref="taxon:10090"
                     /clone="H3099A08"
                     /clone_lib "NIA Mouse 15K cDNA Clone Set"
                     /sex "Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo cDNA libraries (including preimplantation stage
                     embryos from unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos, extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad) and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with Oligo(dT
                     )-Not primers. References include: (1) Genome-wide
                     expression profiling of mid-gestation placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray, 2000,
                     Proc. Natl. Acad. Sci. U S A. 97: 9127-9132; (2)
                     Large-scale cDNA analysis reveals phased gene expression
                     patterns during preimplantation mouse development, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day mouse embryos reveals enrichment in the t-complex
                     and under-representation on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT      170 a    124 c    167 g    133 t
ORIGIN

  Query Match             9.2%;  Score 131.4;  DB 12;  Length 594;
  Best Local Similarity   79.2%;  Pred. No. 6e-26;
  Matches  156;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy        319 CACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAACAAGCATTCAAGGA 378
              || |||| ||||||| || |||| || |||| || |||| ||||||||||||||||
Db          1 CATTATGGGGTCTGGTCCTGCGAAGGCTGCAAGGCTTTCTTTAAGAGAAGCATTCAAGGA 60

Qy        379 CATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAG 438
              || |||| || || ||||||||||||| || || || || || || |||||| | |||
Db         61 CACAATGACTACATGTGTCCAGCTACAAACCAATGCACCATTGACAAGAACCGGAGGAAG 120

Qy        439 AGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTGC 498
              || |||||||||| || || || ||||||||||||||||| ||| |||| | |||
Db        121 AGTTGCCAGGCCTGTCGGCTGCGGCAAGTGTTACGAAGTGGGCATGATGAAAGCGGCATA 180

Qy        499 CGGAGAGAGAGATGTGG 515
              |||| |||  |  | ||
Db        181 CGGAAAGACCGCCGAGG 197


RESULT 11
BM810012
LOCUS       BM810012                1142 bp    mRNA    linear   EST 05-MAR-2002
```

```
DEFINITION  AGENCOURT 6581496 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454957
            5', mRNA sequence.
ACCESSION   BM810012
VERSION     BM810012.1  GI:19126835
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1142)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1950  row: o  column: 22
            High quality sequence start: 13
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .1142
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454957"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      191 a    334 c    398 g    215 t      4 others
ORIGIN

  Query Match             9.1%;  Score 129.8;  DB 14;  Length 1142;
  Best Local Similarity   48.7%;  Pred. No. 2.4e-25;
  Matches  481;  Conservative    0;  Mismatches  499;  Indels    3;  Gaps    1;

Qy        282 CTTCTGCGCCTGTCTGCAGCGATTACGCATCGGGATATCACTATGCAGTCTGGTCGTCTGTGA 341
              |  |||||   |||||  | ||    || || || || |||||||| ||    || |||||
Db         44 CCTCTTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGTGA 103

Qy        342 AGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGC 401
              |  || || ||||| || || || | ||| || ||    |  || || |  ||||| ||
Db        104 GGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGCGGAGCATCGAGTACAGCTGTCCGGC 163

Qy        402 TACAAATCAGTGTACAATCGATAAAAACCGGGCGCAAGAGCTGCCAGCCTGCCGACTTCG 461
                | ||  |||||   |||  ||     | ||||||  ||||||||||||||||  |
Db        164 CTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCAC 223

Qy        462 GAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGTACCG 521
              |||||       ||||| ||| | |||   ||    ||    ||  |      ||    |
Db        224 CAACTGCCTGCCGCTGGGCATGTTCAAGGAGGCAGTGGGCCTGGACGGCGTCCGGGGTGG 283

Qy        522 CCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAG 581
              |   | |    || | | | || | | | | |       |       ||
Db        284 GCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCCC 343

Qy        582 AACTGGCGGCCACGGCCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCA 641
              |||  || ||   |  |  |  | | |  |  | ||||   | ||
Db        344 TGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACT 403

Qy        642 GCTAGTGCTCACCCTGCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGC 701
```

```
              | |   |    || ||| |     |  |        ||      ||||   |
Db        404 GGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGG 463

Qy        702 GCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGT 761
              ||        |  | |    ||   ||||     |  |   |||  ||| | ||
Db        464 CCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATTGT 523

Qy        762 ACACATGATCAGCTGGGCCAAGAAGATCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCA 821
              ||  |||||||||||||||  || || |||||      |||    ||||  |||||
Db        524 GGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCA 583

Qy        822 AGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTC 881
              ||    |  || |||||   ||||||||||||| |  || ||||  ||     ||||||
Db        584 GATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGCTC 643

Qy        882 AATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGG 941
              | |    |  |  |   |||| ||| |||   ||  | || |||||    || |||||
Db        644 ACTGCCACTGCAGGATGAGCTTGGCCTTCGCTGAGGACTTAGTCCTGGA---TGAAGAGGC 700

Qy        942 GAAATGCGTAGAAGGAATTCTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTT 1001
              | | | | ||  ||  |  |||| |    |     || |||  |          || |
Db        701 GGCACGGGCAGCTGGCCTGNGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGCGGCT 760

Qy       1002 TCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAA 1061
              | |  |    || | |  |||| ||| |     | |||||  ||  |||   |||
Db        761 GCAGGCCCTGCGGCTGGAGCGAGAAGAGTATGTTCTACTANAGGCCTTGGCCCTTGCCAA 820

Qy       1062 TTC 1064
              |||
Db        821 TTC 823


RESULT 12
AF455798
LOCUS       AF455798                 137 bp    DNA     linear   GSS 23-APR-2002
DEFINITION  AF455798 Macaca mulatta DNA (Murphy WJ) Macaca mulatta genomic
            clone ESR2, DNA sequence.
ACCESSION   AF455798
VERSION     AF455798.1  GI:20271896
KEYWORDS    GSS.
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 137)
  AUTHORS   Murphy,W.J., Page,J.E., Smith,C. Jr., Desrosiers,R.C. and O'Brien
            ,S.J.
  TITLE     A radiation hybrid mapping panel for the rhesus macaque
  JOURNAL   J. Hered. 92 (6), 516-519 (2001)
  MEDLINE   21945527
COMMENT     Contact: Murphy WJ
            Laboratory of Genomic Diversity
            National Cancer Institute-FCRDC
            P.O. Box B, Bldg. 560, Rm.11-10, Frederick, MD 21702, USA
            Tel: 301 846 1299
            Fax: 301 846 6327
            Email: murphywi@mail.ncifcrf.gov
            Class: unknown.
FEATURES             Location/Qualifiers
     source          1. .137
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /clone="ESR2"
                     /clone_lib="Macaca mulatta DNA (Murphy WJ)"
BASE COUNT       34 a     40 c     30 g     33 t
ORIGIN

  Query Match             8.8%;  Score 126.4;  DB 17;  Length 137;
  Best Local Similarity   95.6%;  Pred. No. 7.2e-25;
  Matches  130;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          5 ATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAA 64
```

```
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          1 ATTACAGCAGTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTGGGCAGACCCCAA 60

Qy         65 GCCCAAATGTGTTGTGGCCAACACCTGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGT 124
              |||||||||||||||||||||||||||||| |||||||||||||||  ||||||||||||
Db         61 GCCCAAATGTGTTGTGGCCAACACCTGGACACCTTTCTCCTTTAGCAGTCCATCGCCAGT 120

Qy        125 TATCACATCTGTATGC 140
              |||||||||| |||||
Db        121 TATCACATCTATATGC 136


RESULT 13
BG891892/c
LOCUS       BG891892                 558 bp    mRNA    linear   EST 04-JUN-2001
DEFINITION  tp46e11.x1 zebrafish gridded kidney Danio rerio cDNA clone 4759773
            3' similar to TR:Q9W669 Q9W669 ESTROGEN RECEPTOR BETA. ;, mRNA
            sequence.
ACCESSION   BG891892
VERSION     BG891892.1  GI:14286502
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,P., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            Genome Systems, St. Louis, Missouri (web address:
            www.genomesystems.com) (email contact: info@genomesystems.com) and
            Research Genetics, Huntsville, Alabama (web address: www.resgen.com
            ) (email contact: info@resgen.com) and
            ResSourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1..558
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="4759773"
                     /clone_lib="zebrafish gridded kidney"
                     /sex="mixed"
                     /tissue_type="kidney pooled from 300 wild type adults"
                     /lab_host="XL01R"
                     /note="Organ: kidney; Vector: pBK CMV; Site_1: EcoRI;
                     Site_2: XhoI; oligo dT cDNA library constructed from mRNA
                     pooled from pooled kidney  tissue from 300 adult
                     zebrafish."
BASE COUNT      137 a    119 c    156 g    146 t
ORIGIN

  Query Match             8.4%;  Score 121;  DB 12;  Length 558;
  Best Local Similarity   68.9%;  Pred. No. 5.2e-23;
  Matches  166;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       1127 TGCTGAACGCCGTGACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGCATCTCCCTCCC 1186
              ||||| || | ||||| || ||| |||||||||  ||  | |||| |||| |   || ||
Db        554 TGCTGGACTCGGTGACTGACGCTCTGGTTTGGGCCATCTCCAAGACCCGCCTTAGCTTCC 495

Qy       1187 AGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGA 1246
              |||||| |||| | ||||| | || || |||| || || ||| |  | || |  |
Db        494 AGCAGCGCTCCACCAGACTGGCCCATCTGCTCATGCTGCTCTCACACATACGCCACGTCA 435

Qy       1247 GTAACAAGGGCATGCAACATCTGCTCAACATGAAGTGCAAAAATGTGGTCCCAGTGTATG 1306
              ||||||| |||||||| || |||| |  |||||||   || ||  |||| ||  ||||||
Db        434 GTAACAAAGGCATGGACCACCTGCACTGCATGAAGATGAAGAAGATGGTTCCTCTGTATG 375

Qy       1307 ACCTGCTGCTGCAGATGCTGAATGCCCACGTGCTTCGCGGGTGCAAGTCCTCCATCACGG 1366
              ||||||| |||||||||||| | || ||| |  | | | | | |   |  ||   | | |
Db        374 ACCTGCTCCTGGAGATGCTGGACGCTCACATCATGCACAGCTCCCGTTTGTCTCACTCCG 315

Qy       1367 G 1367
              |
Db        314 G 314


RESULT 14
BJ068905
LOCUS       BJ068905                 692 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ068905 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL051b20 5', mRNA sequence.
ACCESSION   BJ068905
VERSION     BJ068905.1  GI:17497655
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..692
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL051b20"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
                     /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoKI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett  and A.M. Zorn.
                     (Wellcome/CRC Institute). "
BASE COUNT      169 a    169 c    182 g    182 t
ORIGIN

  Query Match             8.2%;  Score 117;  DB 12;  Length 692;
  Best Local Similarity   51.6%;  Pred. No. 7.9e-22;
  Matches  329;  Conservative    0;  Mismatches  290;  Indels   18;  Gaps    2;

Qy        305 ACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAA 364
              | ||||| || |||||||||||| ||    || |||||  | ||| || ||||| || ||
Db         10 AGGCATCAGGCTATCACTATGGGGTTGCCTCATGTGAGGCATGCAAAGCCTTCTTCAAGC 69

Qy        365 GAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATA 424
              | | ||| |||||  |||  || || |  |||||||| |  ||| | |||  |||    |
Db         70 GCACCATCCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATGAATGTGAAATTACCA 129

Qy        425 AAAACCGGCGCAACAGCTGCCAGGCCTGCGACTTCGGAAGTGTTACGAAGTGGGAATGG 484
```

```
                 | | || || ||| ||| || || ||||| | ||||| |||||| |||| |||
Db        140 AGAGACGTCGAAAGGCCTGTCAAGCTTGCCGCTTCACAAAGTGCCTTCGCGTGGGCATGC 199

Qy        485 TGAAGTGTGGCTCCCGGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTG 544
                 | ||    ||    ||    ||  |     |     | | | |      |||| ||
Db        200 TAAAAGAAGGGGTACGCTTGCACCGTGTCAGAGGAGGACGTCAGAAATACAAGAGGAGGC 259

Qy        545 CCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAG 604
                 | || |   |  | | |                 |     ||| || || || |  | ||
Db        260 CAGATGGAGATATCCTC       ---     CAGTACTCCTCTGCTGCCCACGCTCACCAAC 297

Qy        605 TGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGACCAGCTAGTGCTCACCCTCCTGGAGG 664
                 |      || |  |    |  |  || | | |   |||  || | | |||| ||
Db        298 CCTCTGCAACAGTGGTCAAGAAGCAAACTCCAGTCAATGCAGTTGTCTCTCACCTGCTGG 357

Qy        665 CTGAGCCCCCCCATGTGCTGATCAGCCGCCCCACTGCCCCCTTCACCGAGCCCTCCATC 723
                 ||      ||  |   | || |       ||      |  |  |  | || |      ||
Db        358 TTGCAGAACCAGACAAGTTGTTTGCTATGCCAGACCCAGCACTTCCTGATGGTGATTTGA 417

Qy        724 -     ATGATGTCCCTGACCAAGTTGGCCGACAACGAGTTGGTACACATGATCAGCTGGG 778
                  ||||   |||||    |  | ||  ||     ||  |  ||   ||| || |||||||
Db        418 AATATATGAGCACCCTGTGTGACCTTGCTGATCGTGAAATTGTCATCATCATTAGCTGGG 477

Qy        779 CCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGA 838
                 | |||| || || || ||        ||   ||| |  |||||  || | |||||| |||
Db        478 CAAAGAATATCCCAGGATTCTCCTCCCTTTCCCTCTCGGACCAGATGAGTCTCTTGCAGA 537

Qy        839 GCTGTTGGATGGAGGTGTTAATGATGGGGCTGATCTGGCGCTCAATTGACCACCCCGGCA 898
                 |    ||||||||||| ||| || | ||  |  | |  |||||  ||    |    |
Db        538 GTGTCTGGATGGAGGTTTTATTGCTTGGAGTTGTTTCCGCTCGCTTCCGTATGAAGATG 597

Qy        899 AGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGA 935
                 || |  | ||||||   ||| |||| || ||    |||
Db        598 AGGTTGTGTTTGCTGAGGATTTTGTGCTTGATGAGGA 634


RESULT 15
AZ596395
LOCUS       AZ596395                 532 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  IM0409D10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0409D10 R, DNA sequence.
ACCESSION   AZ596395
VERSION     AZ596395.1  GI:11718585
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Dunn,D., Aoyaqi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Lonqacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tinqey,A., von Niederhausern,A.
            and Wriqht,D.,Weiss,R.
  TITLE     Mouse whole qenome scaffoldinq with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldq., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@qenetics.utah.edu
            Insert Lenqth: 10000  Std Error: 0.00
            Plate: 0409  row: D  column: 10
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 532.
FEATURES             Location/Qualifiers:
     source          1..532
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0409D10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10 Gold, Tl-resistant, F'"
                     /note="Vector: PWD42nv; Purified qenomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.orq/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passaqe throuqh a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oliqonucleotides were
                     liqated to the blunt ends in hiqh  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb ranqe usinq preparative aqarose qel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (qi|4732134|qb|AF129072.1), a copy number
                     inducible derivative of plasmid R1. The vector was liqated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically competent E. coli XL10 Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      118 a    132 c    135 q    147 t
ORIGIN

  Query Match             8.1%;  Score 115.6;  DB 17;  Length 532;
  Best Local Similarity   84.4%;  Pred. No. 1.7e-21;
  Matches  130;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        930 CAGGGATGAGGGCAAATGCGTAGAAGCAATTCTGGAAATCTTTGACATGCTCCTGGCAAC 989
                 ||||||||||||||| |||| |||| ||||||||||||||||||||||||||||||| ||
Db        257 CAGGGATGAGGGGAAGTGCGTGGAAGGCATTCTGGAAATCTTTGACATGCTCCTGGCGAC 316

Qy        990 TACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATATCTCTGTGTCAAGCCCAT 1049
                  ||  || |||| || |||||||||| || ||||||||||||| |||| |||||||
Db        317 GACGGCACGGTTCCGTGAGTTAAAACTGCAGCACAAAGAATATCTGTGTGTGAAGGCCAT 376

Qy       1050 GATCCTGCTCAATTCCAGTATGTACCCTCTGGTC 1083
                 ||| || |||| ||||||  ||| || |  | |
Db        377 GATTCTCCTCAACTCCAGTGAGTAGCCACGAGCC 410


RESULT 16
AQ719425
LOCUS       AQ719425                 611 bp    DNA     linear   GSS 14-JUL 1999
DEFINITION  HS_5512_B2_D07_T7A RPCI-11 Human Male BAC Library Homo sapiens
            qenomic clone Plate=1088 Col=14 Row=H, DNA sequence.
ACCESSION   AQ719425
VERSION     AQ719425.1  GI:5479094
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlonq,J., Younq,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-taqqed connectors: A sequence approach to mappinq and
            scanninq the human qenome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencinq Center
            University of Washinqton
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616 3887
            Email: jwallace@u.washinqton.edu
            Clones are derived from the human BAC library RPCI 11. For BAC
            library availability, please contact Pieter de Jonq
```

```
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Research Genetics (info@resgen.com). BAC end Web Server,
            http://www.htgc.washington.edu
            Plate: 1088  row: H  column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 611.
FEATURES             Location/Qualifiers
     source          1..611
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1088 Col=14 Row=H"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note "Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       178 a    141 c    129 g    144 t     19 others
ORIGIN

  Query Match             8.0%;  Score 115.2;  DB 17;  Length 611;
  Best Local Similarity   97.5%;  Pred. No. 2.4e-21;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        375 AGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCG 434
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 AAGACATAATGATGATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCG 396

Qy        435 CAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 494
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        397 CAAGAGCTGCCAGGCCTGCCGACTTCTGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 456


RESULT 17
BC008218
LOCUS       BC008218                1443 bp    mRNA    linear   HTC 12-JUL-2001
DEFINITION  Homo sapiens, clone IMAGE:3871297, mRNA.
ACCESSION   BC008218
VERSION     BC008218.1  GI:14713363
KEYWORDS    HTC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1443)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, A., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 13 Row: k Column: 10
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 3345517
            This clone has the following problems: Incomplete processing.
FEATURES             Location/Qualifiers
     source          1..1443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3871297"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_67"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT       390 a    339 c    352 g    362 t
ORIGIN

  Query Match             7.7%;  Score 110.6;  DB 11;  Length 1443;
  Best Local Similarity   49.4%;  Pred. No. 7.6e-20;
  Matches  342;  Conservative    0;  Mismatches  324;  Indels   27;  Gaps    1;

Qy        292 GTCTGCAGCGATTACGCATCGGGATATCACTATGCAGTCTGGTCGTGTGAAGGATGTAAG 351
              || ||  | ||   ||| || || || ||||||| ||    || |||||||  || |||
Db        732 GTGTGTGGTGACATCGCTTCTGGGTACCACTATGGGCTAGCATCATGTGAAGCCTGCAAG 791

Qy        352 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 411
              || || || || || |  |||||||  |||  || || |  || || || || ||| |
Db        792 GCATTCTTCAACAGGACAATTCAAGCCAATATACAATACAGCTGCCCTCCCACAAATGAA 851

Qy        412 TGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTAC 471
              |||  ||||   ||   | | || ||   |||||||||| ||||  |   ||||||||
Db        852 TGTGAAATCACAAAGCGCAGACGTAAATCCTGCCAGGCTTGCCGCTTCATGAAGTGTTTA 911

Qy        472 GAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGACAGACAGTGGGTACCGCCTTGTGCGC 531
              |||||| ||| ||||    ||    ||    || |||
Db        912 AAAGTGGGCATGCTGAAAGAAGGGGTGCGTCTTGACAGA---------------------- 950

Qy        532 AGACAGAAGTGCCGACAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGC 591
                  | |||  |    ||||  | ||     | | |  |    |   |  ||
Db        951 ------GTACGTGGAGGTCGGCAGAAGTACAAGGGCAGGATAGATGGGGAGAACAGCCCA 1004

Qy        592 CACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTC 651
              ||  |  ||    || ||  | ||      |   |          | ||    |   |
Db       1005 TACCTGAACCCTCAGCTGGTTCAGCCAGCCAAAAAGCCATATAACAAGATTGTCTCACAT 1064

Qy        652 ACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACC 711
                | |||  |    |  |                ||      ||| | ||  |||  ||
Db       1065 TTGTTGGTGGCTGAACCGGAGAAGATCTATGCCATGCCTGACCCTACTGTCCCCGACAGT 1124

Qy        712 GAGGGCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGA1C 771
              ||   |       | |   | |||    | |||||||||   |||||||     || ||
Db       1125 GACATCAAAGCCCTCACTACACTGTGTGACTTGGCCGACCGAGAGTTGGTGGTTATCATT 1184

Qy        772 AGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTC 831
              | ||||| ||| | ||||| |||||     |||  ||||   ||||| || | ||
Db       1185 GGATGGGCGAAGCATATTCCAGGCTTCTCCACGCTGTCCCTGGCGGACCAGATGAGCCTT 1244

Qy        832 TTGGAGAGCTGTTGCATGGAGGGGTAATGATGGGGCTGATGTGGCGCTCAATTGACCAC 891
              || ||||   |||||||| | || ||  | ||  |  | | |  || ||  ||
Db       1245 CTGCAGAGTGCTTGGATGGAAATTTTGATCCTTGGTGTCGTATACCGGTCTCTTCGTTT 1304

Qy        892 CCCCGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTA 951
              |   | ||  ||| |||    |||  | |  |||||   ||  ||   |||| | |
Db       1305 GAGGATGAACTTGTCTATGCAGACGATTATATAATGGACGAAGACCAGTCCAAATTAGCA 1364

Qy        952 GAAGGAATTCTGGAAATCTTTGACATGCTCCTG 984
              |     ||    ||    ||  || || | |
Db       1365 GGCCTTCTTGATCTAAATAATGCTATCCTGCAG 1397


RESULT 18
BJ069238
LOCUS       BJ069238                610 bp    mRNA    linear   EST 11-DEC 2001
DEFINITION  BJ069238 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL052d06 5', mRNA sequence.
ACCESSION   BJ069238
```

```
VERSION     BJ069238.1  GI:17494935
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559 81 6856
            Fax: 81 559 81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..610
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL052d06"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type "whole embryo"
                     /dev_stage "stage 25"
                     /note "Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett  and A.M. Zorn,
                     (Wellcome/CRC Institute)  "
BASE COUNT      149 a    144 c    165 g    151 t      1 others
ORIGIN

  Query Match             7.6%;  Score 108.8;  DB 13;  Length 610;
  Best Local Similarity   51.7%;  Pred. No. 1.6e-19;
  Matches  410;  Conservative    0;  Mismatches  272;  Indels   18;  Gaps    2;

Qy        292 GTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGTGAAGGATGTAAG 351
              || ||  | |||   ||||| || ||||||||||| ||    || ||||| | ||| ||
Db         14 GTGTCTGGAGATGTGGCATCAGGCTATCACTATGGGGTTGCCTCATGTCAGGCATGCAAA 73

Qy        352 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 411
              ||||| || ||  | | ||| |||||  ||  || || |  |||||||| |  ||| |
Db         74 GCCTTCTTCAAGCGCACCATCCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATGAA 133

Qy        412 TGTACAATCGATAAAAACCGGCGCCAAGACGTCCCAGCCCTGCCGACTTCGGAAGTGTTAC 471
              |||  |||    || |  || || |||  ||| || || ||||||  |    |||||
Db        134 TGTGAAATTACCAAGAGACGTCGAAAGGCCTGTCAAGCTTGCCGCTTCACAAAGTGCCTT 193

Qy        472 GAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGG 531
              ;|||| ||| | ||    ||    ||   ||  |      |     | | |  |
Db        194 CGCGTGGGCATGCTAAAAGAAGGGGTACGCTTGGACCGTGTCAGAGGAGGACGTCAGAAA 253

Qy        532 ACACACAGAAGTCCCGACGAGCAGCTGCACTCTGCCCGCAAGCCCAAGACGAAGTGGCGGT 591
              |||| ||  | || |   | | | |                    |   ||| ||
Db        254 TACAAGAGGAGGCCAGATGGAGATATCCTC-----   ----CAGTACTCGTCTGGTGGG 301

Qy        592 CACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTC 651
              || || |  | ||     |      || |  |    |  |  || | | |  ||| ||
Db        302 CAGGCTCAGCAAGCCTCTGCAACAGTGGTCAAGAAGCAAACTCCAGTGAATGCAGTTGTC 361

Qy        652 ACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACC 711
              | | |||  || ||     || |  | || |      ||     | | | |  |
Db        362 TCTCACCTGCTGGTTGCAGAACCAGACAAGTTGTTTGCTATGCCAGACCCAGCACTTCCT 421

Qy        712 GAGCCCTCCATG-     ATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACAC 765
              || |      ||      ||||  |||||    | | || ||    || | ||   |
Db        422 GATGGTGATTTGAAATATATGAGCACCCTGTGTGACCTTGCTGATCGTGAAATTGTCATC 481

Qy        766 ATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTG 825
```

```
              || || |||||||| |||||  || || ||  ||       ||   ||| | |||||  ||
Db        482 ATCATTAGCTGGGCAAAGAATATCCCAGGATTCTCCTCCCTTTCCCTCTCGGACCAGATG 541

Qy        826 CGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATT 885
              | ||||| ||||    ||||||||||| ||| || | | ||  |  | |  |||||  ||
Db        542 AGTCTCTTGCAGAGTGTCTGGATGGAGGTTTTATTGCTTGGAGTTGTTTCCGCTCGCTT 601


RESULT 19
BG744031
LOCUS       BG744031                 703 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602722811F1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849155 5',
            mRNA sequence.
ACCESSION   BG744031
VERSION     BG744031.1  GI:14054684
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1688  row: f  column: 04
            High quality sequence stop: 628.
FEATURES             Location/Qualifiers
     source          1..703
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4849155"
                     /clone_lib "NIH_MGC_106"
                     /tissue_type="natural killer cells, cell line"
                     /lab_host "DH10B (phage-resistant)"
                     /note="organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      116 a    210 c    245 g    132 t
ORIGIN

  Query Match             7.5%;  Score 107.8;  DB 12;  Length 703;
  Best Local Similarity   48.6%;  Pred. No. 3.2e-19;
  Matches  295;  Conservative    0;  Mismatches  312;  Indels    0;  Gaps    0;

Qy        280 CACTTCTGCGGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
              | | |||||   |||||  | ||    || || || || ||||||| ||    || |||
Db         37 CGCCTCTGGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 96

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db         97 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCG 156

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGACGTCCCAGCCCTGCCGACTT 459
              ||  | ||  |||||   |||   ||     | ||||||  ||||||||||||||||  |
Db        157 GCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 216

Qy        460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTAC 519
                  |||||          ||||| ||| | |||   ||    ||    ||  |     ||
Db        217 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCTACCGGGGT 276
```

```
Qy        520 CGGCTTGTGCGGAGACAGAGAAGACGCCCAGAGCAGCTCTCACCCCCCCCCAAGCCCAAC 579
              | |    | |    || |  |  | || | | | |  |       |          ||
Db        277 GGGCGGCAGAAGTACAAGCCCCCCCGGAGCTCGACCCACTCTCCCTTCCCCCCCCCTC 336

Qy        580 ACAAGTGGGCGCCACCCCCCCCCAGTGCCCCAGCCGCCTGTGCACCCCCTGACCCCCCAC 639
                 |||   ||  ||   |   |   |    | | |   |  ||||         |
Db        337 CCTGCTGGGCCCCTGGCAGTGCTGGAGCCCCCCGAAGACACAGCAGCCCCAGTGAATGCA 396

Qy        640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGCCTGATCAGCCGCCCCAGT 699
              | | |    |    ||  |||  |       |   |         ||      ||||
Db        397 CTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCA 456

Qy        700 GCGCCCTTCACCGAGGCCTCCATGATCATGTCCCTGACCAAGTTGGCCGCAGAAGGAGTTG 759
              |  ||       |  | |    ||   ||||     |   |    |||   |||  | |
Db        457 GGCCCTGATGGGCACCTCCCAGCCGTGGCTATCCTCTGTGACCTCTTTGACGAGACATT 516

Qy        760 GTACACATGATCAGCTGGGCCAAGAAGATCCCGGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
              ||   ||  ||||||||||||||||  || || |||||      |||    ||||  |||
Db        517 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGGTGTCCCTGTCTGAC 576

Qy        820 CAAGTCGGGCTCTTGGAGAGCTGTTGGATGGAGCTGTTAATGATGGGGCTGATGTGGCGC 879
              ||  ||    |  || |||||   ||||||||||||  |  || ||||  ||    |||
Db        577 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCCAGGCCA 636

Qy        880 TCAATTG 886
              |  | ||
Db        637 TGCACTG 643


RESULT 20
BF510841/c
LOCUS       BF510841                 575 bp    mRNA    linear   EST 06-DEC-2000
DEFINITION  UI-H-BI4-apo-e-04-0-U1.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
            IMAGE:3088206 3', mRNA sequence.
ACCESSION   BF510841
VERSION     BF510841.1  GI:11594139
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaphs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone Distribution.
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html The following repetitive
            elements were found in this cDNA sequence: 1-48,
            >AT_rich#low_complexity 349-568, >LINE2
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3088206"
                     /clone_lib="NCI_CGAP_Sub8"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
                     is a subtracted library derived from NCI_CGAP_Sub5. The
                     NCI_CGAP_Sub8 library had 2.5 million recombinants. A
                     single stranded DNA preparation of NCI_CGAP_Sub5 was used
                     as a tracer in a subtractive hybridization with a driver
                     comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE
                     clone Ids 2732833-2737415, 3068840-3069191; 25% of the
                     driver population), a pool of clones from NCI_CGAP_Sub4
                     (IMAGE clone Ids 2723592-2729826; 25% of the driver
                     population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids
                     2728959-2734190; 25% of the driver population), and
                     NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238, 3081864-3084550
                     ; 25% of the driver population). Subtraction was
                     performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To Facilitate Gene Discovery. Genome Research
                     6, 791-806.
                     TAG_LIB=NCI_CGAP_GC4
                     TAG_TISSUE=germ cell
                     TAG_SEQ=AAATC"
BASE COUNT      126 a    131 c    115 g    203 t
ORIGIN

  Query Match             7.5%;  Score 107.4;  DB 12;  Length 575;
  Best Local Similarity   70.2%;  Pred. No. 3.8e-19;
  Matches  144;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         10 AGCATTCCCAGCAAAGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCA 69
              || ||||   |  |||  |   |    | |||   |  ||    | |||   ||
Db        447 AGAATTCAAGGAGATGAGAAGGAAACAGGAGGCAAGATTGAAAAGAAGAGTCCAGTGTGT 388

Qy         70 AATGTGTTGGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCA 129
                ||| |      || |       ||  | ||       ||||||||||||||||||||
Db        387 TATGAGGAAAACCCCAAGAGCATGCTGCCTTACAAGACAGTGGTCCATCGCCAGTTATCA 328

Qy        130 CATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACC 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 CATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACC 268

Qy        190 TTACCTGTAAACAGAGAGACACTGA 214
              |||||||||||||||||||||||||
Db        267 TTACCTGTAAACAGAGAGACACTGA 243


RESULT 21
AK019222
LOCUS       AK019222                 2199 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus 11 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2700080M08:estrogen related receptor,
            alpha, full insert sequence.
ACCESSION   AK019222
VERSION     AK019222.1  GI:12859315
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2700080M08.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,F., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
```

```
                Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
                Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J.,
                Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE         RIKEN integrated sequence analysis (RISA) system--384-format
                sequencing pipeline with 384 multicapillary sequencer
  JOURNAL       Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE       20530913
   PUBMED       11076861
REFERENCE       4
  AUTHORS       Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
                Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
                Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
                Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
                Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
                Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
                Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
                Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
                Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
                Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
                Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
                Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
                Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
                Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
                Ring,B., Ringwald,M., Rodriquez,I., Sakamoto,N., Sasaki,H.,
                Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
                Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
                Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
                and Hayashizaki,Y.
  TITLE         Functional annotation of a full-length mouse cDNA collection
  JOURNAL       Nature 409 (6821), 685-690 (2001)
  MEDLINE       21085660
   PUBMED       11217851
REFERENCE       5  (bases 1 to 2199)
  AUTHORS       Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
                Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
                Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
                Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
                Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
                Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
                Kuribara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
                Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
                Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
                Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
                Sogabe,Y., Suzuki,H., Taqami,M., Taqawa,A., Takahashi,F.,
                Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
                Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
                Hayashizaki,Y.
  TITLE         Direct Submission
  JOURNAL       Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
                Physical and Chemical Research (RIKEN), Laboratory for Genome
                Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
                RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
                Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
                URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503 9222,
                Fax:81 45-503-9216)
COMMENT         Please visit our web site (http://genome.gsc.riken.go.jp/) for
                further details.
                cDNA library was prepared and sequenced in Mouse Genome
                Encyclopedia Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                Division of Experimental Animal Research in Riken contributed to
                prepare mouse tissues. First strand cDNA was primed with a primer
                [5' GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                prepared by using trehalose thermo-activated reverse transcriptase
                and subsequently enriched for full length by cap-trapper. Second
                strand cDNA was prepared with the primer adapter of sequence [5'
                GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
                with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
                Host: SOLR.
FEATURES                Location/Qualifiers
     source             1. .2199
                        /organism "Mus musculus"
                        /strain "C57BL/6J"
                        /db_xref "FANTOM_DB:2700080M08"
                        /db_xref "MGD:MGI:1899528"
                        /db_xref "taxon:10090"
                        /clone "2700080M08"
                        /clone_lib "RIKEN full length enriched mouse cDNA library"
                        /dev_stage="11 days embryo"
     gene               1. .2199
                        /gene="Esrra"
     misc_feature       1. .2199
                        /gene "Esrra"
                        /note "Data Source:MGD, source key:MGI:1346831,
                        evidence:ISS
                        estrogen related receptor, alpha"
                        /db_xref="MGD:MGI:1346831"
BASE COUNT      425 a    622 c    722 g    430 t
ORIGIN

  Query Match             7.5%;  Score 107.4;  DB 11;  Length 2199;
  Best Local Similarity   48.7%;  Pred. No. 7.7e-19;
  Matches  322;  Conservative    0;  Mismatches  336;  Indels    4;  Gaps    1;

Qy        280 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
              | | |||||   ||||| | |||   || || || || ||||| || ||    || |||
Db        389 CGCCTCTGCCTGGTCTGTGGGGATGTGGCCTCTGGCTACCACTACGGTGTGGCATCCTGT 448

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db        449 GACGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGAGCATCGAGTACAGCTGTCCG 508

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||| |||||   |||   ||      | ||||||  ||| ||||||||||  |
Db        509 GCCTCCAATGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGTCAGGCCTGCCGCTTC 568

Qy        460 CGCAAGTGTTACCAACTGGCAATGGTGAAGTGTGGCCTCCCGGAGAGAGATGTGGGTAC 519
              |||||       ||||| ||| | |||   ||    ||    ||     |||   |
Db        569 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGTGTGCGTCTGGA---CCGTGTCCGC 625

Qy        520 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 579
               ||    || ||   | | | |      ||| |  |  |  |  |  | || |
Db        626 GGCGGACGGCAGAAGTACAAACGGCGGCCAGAGGTGGACCCTTTGCCTTTCCCGGGCCCC 685

Qy        580 ACAAGTGGCGGCCACCGCGCCCGAGTGCGGGCAGCTGCTGCTGGACGCCCTGACCCCGAG 639
                 ||  || |    | |  ||  | |  |  |  |  |||    ||
Db        686 TTCCCTGCTGGACCTCTGGCAGTAGCTGGAGGACCCAGGAAGACAGCCCCAGTGAACGCT 745

Qy        640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 699
              | | |  ||    ||  ||| |    |  |    ||    |||
Db        746 CTGGTGTCGCATCTGCTGGTGGTTGAACCTGAGAAGCTGTACGCCATGCCTGACCCAGCA 805

Qy        700 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 759
               |||      |  | ||    ||  | ||    | |  ||    ||  |||
Db        806 AGCCCCGATGGACACCTCCCCGCTGTGGCCACTCTCTGTGACCTTTTTGATCGAGAGATA 865

Qy        760 GTACATGATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
              ||  ||  |||||||||||||||  || || ||||    ||    |||| |||
Db        866 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCCTCACTGTCACTGTCTGAC 925

Qy        820 CAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGC 879
              ||  ||    |  || ||||    |||||||| ||| |  || ||||  ||      |||
Db        926 CACATCTCACTACTGCCACAGTGTGTGGATGGAAGTGCTGGTGCTGGCTCTGGCCCACCGC 985

Qy        880 TCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAG 939
              ||| |    |  | |  ||||   |||||||   || || ||  |  |    || | |
Db        986 TCACTGCCACTGCAGGATGAGCTGGCCTTTGCTGAGGACCTGGTCTTATATGAAGAGGGG 1045

Qy        940 G 940
              |
Db       1046 G 1046


RESULT 22
BG747925
ID   BG747925                 676 bp    mRNA    linear   EST 15-MAY 2001
```

```
DEFINITION  602705778F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842402 5'
            mRNA sequence.
ACCESSION   BG747925
VERSION     BG747925.1  GI:14058578
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1676  row: 1  column: 19
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1..676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842402"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies)
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      113 a    209 c    222 g    132 t
ORIGIN

  Query Match             7.4%;  Score 106.2;  DB 12;  Length 676;
  Best Local Similarity   48.7%;  Pred. No. 9e-19;
  Matches  288;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;

Qy        280 CACTTCTGCGGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
              | | |||||   |||||  | ||    || || || || ||||||||  ||      || |||
Db         26 CGGCTCTGCCTGGTCTGTGGGGACGTGGCCTCGGGCTACCACTATCGTGTCGCCATCGTGT 85

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGCACATAATGATTATATTTGTCCA 399
              || |  || || ||||| || || || | ||| || ||    |  || || |  |||||
Db         86 GAGGCCTGCAAAGCCTTCTTCAAGAGCACCATCCAGGGGAGCATCGAGTACAGCTGTCCC 145

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACGCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||  |||||   |||   ||     | ||||||  ||||||||||||||  |
Db        146 GCCTCCAACGAGTGTGAGATCACCAAGCGGCAGACGCAAGGCCTGCCAGGCCTGCCCCTTC 205

Qy        460 CGGAAGTGTTACCAAGTGGGAATGGTGAAGTGTGGCTCCCCGCAGACAGACATCTCCGTAC 519
              |||||       ||||| ||| | |||   ||    ||    ||  |      ||
Db        206 ACCAAGTCCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCAGTCCTGCAAGCACTCCATCGAC 265

Qy        520 CGCGTTGTGCGCAGACAGAGAAGTGCCGAAGAGCAGCTGCACTGTGGCCGGCAAGCCAAG 579
              | |    | |   || |  | | || | | | |  |       |          ||
Db        266 GGGCGGCAGAACTACAAGGGCGTGCCGGAGTGGACCAAGTGCCTTCAGCCGCCGCCTTC 325

Qy        580 AGAAGTGGCGGCCAGGCGCCCGAGTGCGGGAGCTGCTGCTGGAAGCCTGAGCCCGAG 634
                 |||   ||  ||   |  |   |    | |  |   |  ||||       |
Db        326 CCTGCTGGGCCCTGGCAGTCGCTGCAGGCCCCGGAAGACAGCAGCCCCAGTGAATGCA 385

Qy        640 CAGCTAGTGCTCACCGTCTGGAGGCTGAGCGGCCCATCTGCTGATCAGCGCCCCAGT 699
              | | |    |     ||  |||  |       |  |       ||      ||||
```

```
Db        386 CTGCTCTCATCTCCTCTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCCA 445

Qy        700 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 759
              |  ||        |   | |     ||    ||||     | |    |||   ||| |
Db        446 GGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGAGAGATT 505

Qy        760 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
              ||   ||  ||||||||||||||||  || || ||||       |||    ||||  |||
Db        506 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 565

Qy        820 CAAGTGGGCTCTGGAGAGTTGTTGGATGGAGCTGTTAATGATGGGGCTG 870
              ||  ||    |  || |||||   ||||||||||||| |  || ||||  ||
Db        566 CAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTG 616


RESULT 23
BI100868
LOCUS       BI100868                844 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602886184F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5041580
            5', mRNA sequence.
ACCESSION   BI100868
VERSION     BI100868.1  GI:14551761
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11114  row: 9  column: 21
            High quality sequence stop: 775.
FEATURES             Location/Qualifiers
     source          1..844
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5041580"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      156 a    244 c    276 g    168 t
ORIGIN

  Query Match             7.1%;  Score 102;  DB 13;  Length 844;
  Best Local Similarity   48.7%;  Pred. No. 1.6e-17;
  Matches  307;  Conservative    0;  Mismatches  320;  Indels    3;  Gaps    1;

Qy        311 CGGCATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCA 370
              |||| || ||||| || ||    || ||||| |  || || ||||| || || || | ||
Db         31 CGGGCTACCACTACGGTGTCGCCATCGTGTGAGGCCTGCAAAGCCTTCTTCAACAGGACCA 70

Qy        371 TTCAAGCACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACC 430
              | || ||    |  || || |  ||||| ||  | ||| ||||   |||   ||
Db         71 TCCAGGGGAGCATCGAGTACAGCTGTCCGGCCTCCAATGAGTGTGAGATCACCAAGCGGA 130

Qy        431 GGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACCAAGTGGGAATGGTGAAGT 490
              | ||||||  ||| ||||||||||  |   |||||       ||||| ||| | |||
Db        131 GACGCAAGGCCTGTCAGGCCTGCCGCTTCACCAAGTGCCTGCGGGTGGGCATGCTCAAGG 190
```

```
Qy        491 GTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACG 550
              ||    ||    ||  |     |  | | |    | |    |  |  | | || |
Db        191 AGGGTGTGCGTCTGGACCGTGTCCGCGGCGGACGGCAGAAGTACAAACGGCGGCCAGAGG 250

Qy        551 AGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGG 610
              | | |       | |      ||         |||    | ||      |  |
Db        251 TGGACCCTTTGCCTTTCCCGGGCCCCTTCCCTGCTGGACCTCTGGCAGTAGCTGGAGGAC 310

Qy        611 AGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGC 670
              | |   | | |  ||| | |   | |     ||     | | ||  |||||
Db        311 CCAGGAAGACAGCCCCAGTGAACGCTCTGGTGTCGCATCTGCTGGTGGTTGAACCTGAGA 370

Qy        671 CGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGT 730
              ||   | |   ||     ||   | ||    |||        |   | ||    ||
Db        371 AGCTGTACGCCATGCCTGACCCAGCAAG---CCCCGATGGACACCTCCCCGCTGTGGCCA 427

Qy        731 CCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTC 790
              | ||     |  |    ||    ||| | ||   ||  ||||||||||||||||  || |
Db        428 CTCTCTGTGACCTTTTTGATCGAGAGATAGTGGTCACCATCAGCTGGGCCAAGAGCATCC 487

Qy        791 CCGGCTTTGTGGCAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGG 850
              | ||||      ||    ||||  |||||  ||    |  || ||||      |||||||
Db        488 CAGGCTTCTCCTCACTGTCACTGTCTGACCAGATGTCAGTACTGCAGAGTGTGTGGATGG 547

Qy        851 AGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTG 910
              | ||| |  || |||   ||     |||||||  |    |  |   |||   |||||
Db        548 AAGTGCTGGTGCTGGGTGTGGCCCAGCGCTCACTGCCACTGCAGGATGAGCTGGCCTTTG 607

Qy        911 CTCCAGATCTTGTTCTGGACAGGGATGAGG 940
              ||   || || || || ||    || | ||
Db        608 CTGAGGACCTGGTCCTAGATGAAGAGGGGG 637



RESULT 24
BM562894
LOCUS       BM562894                 937 bp    mRNA    linear   EST 20 FEB 2002
DEFINITION  AGENCOURT_6588670 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478584
            5', mRNA sequence.
ACCESSION   BM562894
VERSION     BM562894.1  GI:18809360
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1998  row: h  column: 09
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1..937
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5478584"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Bonq
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      182 a    286 c    297 g    170 t      2 others
ORIGIN

  Query Match           7.1%;  Score 101.4;  DB 13;  Length 937;
  Best Local Similarity   47.9%;  Pred. No. 2.4e-17;
  Matches  291;  Conservative    0;  Mismatches  316;  Indels    0;  Gaps    0;

Qy        334 TCGTGTGAAGGATGTAAGGCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATT 393
              || ||||| |  || || ||||| || || || | ||| || ||    |  || || |
Db          5 TCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGC 64

Qy        394 TGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCTGC 453
              ||||| ||  | ||  ||||| |||  ||     | |||||| ||||||||||||
Db         65 TGTCCGGCCTCCAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGC 124

Qy        454 CGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGT 513
              ||  |   |||||        ||||| ||| | |||   ||    ||    ||  |
Db        125 CGCTTCACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTC 184

Qy        514 GGGTACCGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAG 573
              ||    | |   |  | |   || |  |  | || | | | |  |       |
Db        185 CGGGGTGGGCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGC 244

Qy        574 GCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGC 633
              ||          |||   ||  ||   |  |  |  |  | | |  |  ||||
Db        245 CCCTTCCCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTG 304

Qy        634 CCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGC 693
                |  | | |    |    ||  |||  |    |   |    ||      |
Db        305 AATGCACTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATGCCTGAC 364

Qy        694 CCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAG 753
              |||   |  ||        |  | |    ||   ||||   |  |   |||
Db        365 CCCGCAGGCCCTGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACCGA 424

Qy        754 GAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTG 813
              ||| | ||   ||  |||||||||||||||  || || ||||      |||   |||
Db        425 GAGATTGTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTG 484

Qy        814 TTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTTAATGATGGGGCTGATG 873
              |  |||||  ||    |  || |||||   |||||||||||| |  || ||||  ||
Db        485 TCTGACCAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTGGCC 544

Qy        874 TGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCTCCAGATCTTGTTCTGGACAGG 933
              |||||||  |    |  |  ||||   ||| |||   ||  | || ||||
Db        545 CAGCGCTCACTGCCACTGCAGGATGAGCTGGCCTTCGCTGAGGACTTACTCCTGGATGAA 604

Qy        934 GATGAGG 940
              || | ||
Db        605 GAGGGGG 611



RESULT 25
BJ058831
LOCUS       BJ058831                 558 bp    mRNA    linear   EST 11 DEC 2001
DEFINITION  BJ058831 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL060114 5', mRNA sequence.
ACCESSION   BJ058831
VERSION     BJ058831.1  GI:17492544
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
```

```
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@qenes.niq.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL060113"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
                     /note="Vector: pBSFN3; Site_1: NotI; Site_2: EcoRI. cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett  and A.M. Zorn,
                     (Wellcome/CRC Institute)  "
BASE COUNT      142 a    134 c    146 g    136 t
ORIGIN

  Query Match             7.0%;  Score 100.4;  DB 13;  Length 558;
  Best Local Similarity   51.6%;  Pred. No. 3.6e-17;
  Matches  292;  Conservative    0,  Mismatches  256,  Indels   18,  Gaps    2;

Qy        307 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAACGATGTAAGGCCTTTTTTAAAACA 366
              ||||| || |||||||||||| ||    || ||||| | ||| || ||||| || || | 
Db          1 GCATCAGGCTATCACTATGGGGTTGCCTCATGTGAGGCATGCAAAGCCTTCTTCAAGCGC 60

Qy        367 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTCTACAATCCATAAA 426
              | ||| ||||| ||| || || | |||||||| | ||| | ||| ||| ||    || 
Db         61 AGCATCCAAGGCAATATCGAGTACAGCTGTCCAGCCAGCAATGAATGTGAAATTACCAAC 120

Qy        427 AACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTG 486
              | || || ||| ||| || || ||||| | ||||| ||||| ||| | 
Db        121 AGACGTCGAAAGGCCTCTCAAGCTTGCCGCTTCACAAAGTGCCTTCGCGTCCGCATCCTA 180

Qy        487 AAGTGTGGCTCCCGGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCC 546
              ||    ||    ||    ||  |     |     | | |  |        |||| ||  | 
Db        181 AAAGAAGGGGTACGCTTGGACCGTGTCAGAGGAGGACGTCAGAAATACAAGAGGAGCCA 240

Qy        547 CACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCAGGCGCCCGAGTG 606
              || |   | | | |                 |     ||| || || || |  | ||  
Db        241 GATGGAGATATCCTC------------CAGTACTCGTCTCGGCAGGCTCAGCAACCC 285

Qy        607 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 666
              |     || |  |    | || | || | | |   |||  || | | ||||  | | | 
Db        289 TCTGCAACAGTGGTCAAGAAGCAAACTCCAGTGAATGCAGTTGTCTCTCACCTGCTGGTT 348

Qy        667 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATG--- 723
              |      ||  |   | || |       ||     |  |  |  | || |       || 
Db        349 GCAGAACCAGACAAGTTGTTTGCTATGCCAGACCCAGCACTTCCTGATGGTCATTTGAAA 408

Qy        724 ---ATGATGTCCCTGACCAAGTTGGCCGACAAGCAGTTGGTACACATGATCAGCTGGGCC 780
                 ||||  |||||    | | || ||    ||  | ||   ||| || ||||||||| 
Db        409 TATATGAGCACCCTGTGTGACCTTGCTGATGTGAAATTGTCAATCATCAAGCTGGGCA 468

Qy        781 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCCACCAAGTCCCGCTCTTGCACAGC 840
              ||||| || || || ||      ||   ||| |  |||||  || |  |||||| |||| 
Db        469 AAGAATATCCCAGGATTCTCCTCCCTTTCCCTCTCGGACCAGATGAGTCTCTTGCAGACT 528

Qy        841 TGTTGGATGGAGGTGTTAATGATGGG 866
              |||||||||||| ||| || | || 
Db        529 GTCTGGATGGAGGTTTTATTGCTTGG 554


RESULT 26
BE546976
LOCUS       BE546976                908 bp    mRNA    linear   EST 09-AUG-2000
DEFINITION  601071950F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3457885 5', 
            mRNA sequence.
ACCESSION   BE546976
VERSION     BE546976.1  GI:9775621
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 908)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8448  row: d  column: 14
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .908
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3457885"
                     /clone_lib "NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      172 a    251 c    334 g    151 t
ORIGIN

  Query Match             7.0%;  Score 99.8;  DB 10;  Length 908;
  Best Local Similarity   48.3%;  Pred. No. 6.8e-17;
  Matches  278,  Conservative    0,  Mismatches  297;  Indels    0;  Gaps    0;

Qy        307 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGA 366
              || || || || |||||||| ||    || ||||| | || || |||| || || || 
Db         11 GCCTCCGGCTACCACTATGGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGC 70

Qy        367 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTCTACAATCCATAAA 426
              | ||| || ||    | || || | ||||| || | ||  ||||   |||  || 
Db         71 ACCATCCAGGGGAGCATCGAGTACAGCTGTCCGGCCTCCAACGAGTGTGAGATCACCAAG 130

Qy        427 AACGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTG 486
              | ||||||  |||||||||||||||| |    |||||       ||||| ||| | 
Db        131 CGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAAGTGCCTGCGGCTGGGCATGCTC 190

Qy        487 AAGTGTGGCTCCCGGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCC 546
              |||   ||    ||    ||  |    ||    | |    |  |    ||  | | | 
Db        191 AAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGTGGGCGGCACAAGTACAAGCGGCGCCG 250

Qy        547 CACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCAGGCGCCCGAGTG 606
              || | | | |  |        |         ||        |||  ||  ||  |  | 
Db        251 CAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCCCTGCTGGGGCCCTCCCACTCGCTGGA 310

Qy        607 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 666
              |     | | |   | ||||         | | | |     |    |||  | 
Db        311 CGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACTGGTGTCTCATCTGCTGGTGGTTGAG 370

Qy        667 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATG 726
              |   |         ||    ||||  | ||        |  | |     ||    || 
Db        371 CCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCCCTGATGGGCACCTCCCAGCCGTG 430

Qy        727 ATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAG 786
                ||||    |  |   |||  ||| | ||   ||  ||||||||||||||||| 
Db        431 CCTACCCTCTGTGACCTCTTTGACCGAGAGATTGTGGTCACCATCAAGCTGGGCCAAGAGC 490
```

```
Qy          787 ATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTTGC 846
                ||  || ||||     ||| |   |||| ||||  ||   |  || ||||   |||
Db          491 ATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAGATGTCAGTACTGCAGAGCGTGTGC 550

Qy          847 ATGGAGGTGTTAATGATGGGCTGATGTGGCGCTC 881
                |||||||||| |  || |||  ||    ||| | |
Db          551 ATGGAGGTGCTGGTGCTGGGTGTGGCCAGGCTCAC 585


RESULT 27
AU169990
LOCUS       AU169990                800 bp    mRNA    linear   EST 29-JAN 2001
DEFINITION  AU169990 ol-br-ad cDNA Oryzias latipes cDNA clone br5464, mRNA
            sequence.
ACCESSION   AU169990
VERSION     AU169990.1  GI:12592059
KEYWORDS    EST.
SOURCE      Japanese medaka.
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Mita,K., Ishikawa,Y. and Yamauchi,M.
  TITLE     Establishment of cDNA database of medaka, Oryzias latipes
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mita K
            Genome Research Group
            National Institute of Radiological Sciences
            Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
            Email: kmita@nirs.go.jp
            method:uni-directional sequence direction:sequenced from T3 primer
            (5' --> 3').
FEATURES             Location/Qualifiers
     source          1..800
                     /organism="Oryzias latipes"
                     /strain "HNI"
                     /db_xref="taxon:8090"
                     /clone "br5464"
                     /clone_lib "ol br-ad cDNA"
                     /sex "female/male mixed"
                     /tissue_type "brain"
                     /dev_stage "adult"
BASE COUNT      198 a    212 c    238 g    152 t
ORIGIN

  Query Match           6.9%;  Score 99.6;  DB 9;  Length 800;
  Best Local Similarity   65.0%;  Pred. No. 7.3e-17;
  Matches  147;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 343
                ||||   || ||  | |||   || || ||  ||||||||||||||   ||| |||||||
Db          568 TCTGCCTCGTATGTGGGGATGTGGCGTCCGGGTATCACTATGGAGTGGCGTCATGTGAAG 627

Qy          344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
                  || || ||  || || ||  |||| ||||||||  | |  ||  || |  ||||||||
Db          628 CCTGCAAAGCTTTCTTCAAGAGAACCATTCAAGGTAACATCGAGTACAGCTGTCCAGCAT 687

Qy          404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
                ||||  ||||   ||||   |||| || | |||   ||||| || ||||| ||   |
Db          688 CAAACGAGTGCGAAATCACCAAAAGACGCACAAAGGCGTGCCAAGCATGCCGCTTTACCA 747

Qy          464 AGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAG 509
                ||||   | |||| || ||| | |||   ||    ||   ||| ||
Db          748 AGTGCCTCAAAGTAGGCATGCTCAAGGAGGGAGTGCGTCTAGATAG 793


RESULT 28
AL527012
LOCUS       AL527012                1004 bp    mRNA    linear   EST 13-FEB 2001
DEFINITION  AL527012 LTI_NFL004_NBC3 Homo sapiens cDNA clone CS0DC017YG10 5
            prime, mRNA sequence.
ACCESSION   AL527012
VERSION     AL527012.1  GI:12790505
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..1004
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "CS0DC017YG10"
                     /clone_lib "LTI_NFL004_NBC3"
                     /sex="male"
                     /tissue_type "neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a Not I oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      163 a    324 c    350 g    164 t      3 others
ORIGIN

  Query Match           6.9%;  Score 99.6;  DB 9;  Length 1004;
  Best Local Similarity   49.0%;  Pred. No. 8.2e-17;
  Matches  290;  Conservative    1;  Mismatches  300;  Indels    1;  Gaps    1;

Qy          280 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
                | | |||||   |||||  | ||    || || || || |||||||| ||    || |||
Db          380 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 439

Qy          340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
                || |  || || |||| || || | ||| || ||    |  || || |  ||||||
Db          440 GAGGCCTGCAAAGCTTCTTCAAGAGGACCATCCAGGGGAGCATCGAGTACAGCTGTCCA 499

Qy          400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
                ||  | ||  |||||   |||  ||     | |||||  |||||||||||||||  |
Db          500 GCCTCCAACGAGTGTGAGATCACCAAGCGGACGCAAGGCTGCCAGGCCTGCCGCTTC 559

Qy          460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGTAC 519
                  |||||       |||||  ||| | |||   ||    ||    ||  |    ||
Db          560 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGG 619

Qy          520 CGCCTTGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAG 579
                |   |       |||      |||| |    |   |||||  |   |  ||||
Db          620 TGGGCGGCAGAAGTACAAGCGGCGGCGGCGAGGTGGGACCCACTGCCCTTCCCGGCCCCT 679

Qy          580 AGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAG 639
                    |||| |  ||      ||  || |     |   |     |          |
Db          680 TCCCTGCTGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATG 739

Qy          640 CAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGT 699
                ||       |||| |  |  |  |      | |         ||    |||   ||
Db          740 CACTGGTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCCTCTATGCCATGCCTGACCCCG 799

Qy          700 GCGCCCTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTG 759
                  |||||     |  | |    ||    ||||      |  |    |||   ||| |
Db          800 CAGCCCTGATGGGCACTCCCAGCCGTGGCTACCCTCTGTGACCTCTTGACCGAGAGATT 859
```

```
Qy        760 GTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGAC 819
              ||   ||  ||||||||||||||||  || |  |||||      |||    ||||  |||
Db        860 GTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGAC 919

Qy        820 CAAGTGCGGCTCTTGGAGAGC-TGTTGCATGGAGCTGTTAATGATGGGGCTG 870
              ||  ||    |  || |||:| || ||||| |||||| |  || ||| | ||
Db        920 CAGATGTCAGTACTGCAGAGCGTGGTGGATTGAGGTGCTGGTGCTGGTGTG 971


RESULT 29
BG480960
LOCUS       BG480960                 775 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602529118F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4652501 5',
            mRNA sequence.
ACCESSION   BG480960
VERSION     BG480960.1  GI:13413239
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1437  row: d  column: 06
            High quality sequence stop: 663
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4652501"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G)  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  SAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      131 a    215 c    282 g    146 t
ORIGIN

  Query Match             6.7%;  Score 96.6;  DB 12;  Length 775;
  Best Local Similarity   51.1%;  Pred. No. 5.1e-16;
  Matches  303;  Conservative    0;  Mismatches  284;  Indels    6;  Gaps    4;

Qy        280 CACTTCTGCGCTGTCTGCCAGCGATACGCATCGCGATATCACTATGGAGTCTGGTCGTGT 339
              | | |||||   |||||  | ||    || || || || |||||||| ||    || |||
Db         83 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTGTGGCATCCTGT 142

Qy        340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
              || |  || || ||||| || || || | ||| || |||   |  || || |  |||||
Db        143 GAGGCCTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGA-GCATCGAGTACAGCTGTCCG 201

Qy        400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
              ||  | ||  |||||   |||   ||      | ||||||  |||||||||||||||  |
Db        202 GCCTCCAACGACTGTGAGATCACCAAGCGGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC 261

Qy        460 CGGAAGTGTTACCAAGTGGGAATGGTGAAG--TGTGGCTCCCGGAGACAGACATGTGGGT 517
```

```
              |||||       |||||  ||| | |||   |  |  | |   ||  | |  | ||
Db        262 ACCAAGTGCCTGCGGCTGGGCATGCTCAAGGAGGGAGTGCGCTTGCACCCGCTGGGGTG 321

Qy        518 ACCGGCTTGTGCCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGCCAAGGCA 577
              ||  |   |  | | | |   | |  || |  |||| ||  | |||     |
Db        322 GCCGGCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCC 381

Qy        578 AGAGAAGTGGCGGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCG 637
                |  |      |   | ||| |   |  |       | | |  |||   |
Db        382 CTGCTGGCCCCCTGGCAGTCGCTGGACGCCCCCGGAAGACAGCAGCCCCAGTGAATGCAC 441

Qy        638 AGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCA 697
              |  |       ||      |  | ||| |||||  || | |||   ||      || | |
Db        442 TGGTGTCTCATCTGCTCGTGGTTGACCCTGAGAAGCTCTATGCCATGCCTGACCCCCCAG 501

Qy        698 CTGCGCCCTTCACCGAGCCCTCCATCATGATCTCCCTGACCAAGTTGCCCGACAAGGACT 757
              |  |         |   | || ||  |  ||||    |  |   |||   |||
Db        502 GCCCTGATGGGCACCCTCCCAGCCGTGGCTA---CCCTCTGTGACCTCTTTGACCGAGAGA 558

Qy        758 TGGTACACATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCG 817
              | ||   ||  ||||||||||||||||  || || |||||      |||    ||||  |
Db        559 TTGTGGTCACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTG 618

Qy        818 ACCAAGTGCGGCTCTTGGAGAGCTGTTGCATGGAGCTGTTAATGATGGGGCTG 870
              ||||  ||    |  || |||||   ||||||||||||| |  || ||||  ||
Db        619 ACCAGATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGTGTG 671


RESULT 30
BG033384
LOCUS       BG033384                 979 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  602381556F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4403197 5',
            mRNA sequence.
ACCESSION   BG033384
VERSION     BG033384.1  GI:12425627
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM18112  row: p  column: 14
            High quality sequence stop: 681.
FEATURES             Location/Qualifiers
     source          1. .979
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4403197"
                     /clone_lib="NIH_MGC_87"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.384 kb. Library enriched for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      150 a    291 c    353 g    184 t      1 others
ORIGIN

  Query Match             6.7%;  Score 96.4;  DB 12;  Length 979;
  Best Local Similarity   49.0%;  Pred. No. 6.5e-16;
```

```
  Matches  285;  Conservative    0;  Mismatches  296;  Indels    1;  Gaps    1;

Qy         307 GCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGA 366
               || || || || |||||||| ||    || ||||| | || || ||||| || || ||
Db           4 GCCTCCGGCTACCACTATGGCTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGG 63

Qy         367 AGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAA 426
               | ||| || ||| |  || || | ||||||||  | ||  ||||| |||   ||
Db          64 ACCATCCAGGGAGC ATGGAGTACAGCTGTCCAGCCTCCAACGAGTGTGAGATCACCAAG 122

Qy         427 AACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGCTG 486
               | ||||||  ||||||||||||||||  |    |||||        ||||| ||| |
Db         123 CGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAAGTGCCTGCGGGTGGGCATGCTC 182

Qy         487 AAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGAAGTGCC 546
               |||  ||   ||   || | |    ||  | |   | |    || | | |  | |
Db         183 AAGGAGGGAGTGCGCCTGGACCGCGTCCGGGGTGGGCGGCAGAAGTACAAGCGGCGGCCG 242

Qy         547 GACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTG 606
               || | | | | |     |    ||      ||| || ||  |   |
Db         243 GAGGTGGACCCACTGCCCTTCCCGGGCTCCTTCCCTGCTGGGCCCCTGGCAGTCGCTGGA 302

Qy         607 CGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCT 666
               |    | |  | |  | ||||       | | | |    |    || ||| |
Db         303 GGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACTGGTGTCTCATCTGCTGGTGGTTGAG 362

Qy         667 GAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATG 726
               |  |     ||      |||| |  ||       |  | |  |     ||
Db         363 CCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCCCTGATGGGCACCTCCCAGCCGTG 422

Qy         727 ATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGCTACACATGATCAGCTGGGCCAAGAAG 786
               ||||    |  |    ||| ||| | ||   ||  ||||||||||||||||
Db         423 GCTACCCTCTGTGACCTCTTTGACCGAGAGATTGTGGTCACCATCAGCTGGGCCAAGAGC 482

Qy         787 ATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGC 846
               || || ||||      |||   ||||  |||||  ||    |  || ||||   |||
Db         483 ATCCCAGGCTTCTCATCGCTGCGCTGTCTGACCAGATGTCAGTACTGCAGAGCGTGTGC 542

Qy         847 ATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGAC 888
               |||||||||  |  || ||||  ||     ||||||| |  ||
Db         543 ATGGAGGTGCTGGTGCTGGGTGTGGCCCAGCGCTCACTGCAC 584


RESULT 41
BI738831
LOCUS       BI738831                 818 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603361587F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5368892 5',
            mRNA sequence.
ACCESSION   BI738831
VERSION     BI738831.1  GI:15715344
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1988  row: m  column: 21
            High quality sequence stop: 817.
FEATURES             Location/Qualifiers
     source          1..818
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5368892"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="organ: eye; Vector: pCMV-SPORT6; Site_1: Not1;
                     Site_2: Sal1; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      185 a    245 c    230 g    158 t
ORIGIN

  Query Match             6.6%;  Score 95;  DB 13;  Length 818;
  Best Local Similarity   65.1%;  Pred. No. 1.5e-15;
  Matches  140;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         280 CACTTCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGT 339
               | | | |||   || ||| | ||    || || || || ||||| |||||     || ||
Db         303 CGCCTGTGCCTCGTGTGCGGGGACATTGCCTCTGGCTACCACTACGGAGTGGCCTCCTGC 362

Qy         340 GAAGGATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCA 399
               || |  || |||| || || || |||| |||||||||  | |  || || | |  || ||
Db         363 GAGGCTTGCAAGGCGTTCTTCAAGAGAACCATTCAAGGCAACATCGAGTACAACTGCCCG 422

Qy         400 GCTACAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTT 459
               || || ||| | |||   |||   |||    ||||||||  ||| ||||||||||||| |
Db         423 GCCACCAATGAATGTGAGATCACCAAACGGAGGCGCAAGTCCTGTCAGGCCTGCCGATTC 482

Qy         460 CGGAAGTGTTACGAAGTGGGAATGGTGAAGTGTGG 494
               ||| ||   | |||||| ||| |||||   ||
Db         483 ATGAAATGCCTCAAAGTGGGGATGCTGAAGGAAGG 517


RESULT 42
BF717864
LOCUS       BF717864                 402 bp    mRNA    linear   EST 07-JUN-2001
DEFINITION  fd52h04.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
            IMAGE:3733497 5' similar to TR:O13012 O13012 ESTROGEN RECEPTOR. ;,
            mRNA sequence.
ACCESSION   BF717864
VERSION     BF717864.1  GI:12016963
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopteryqii; Neopteryqii; Teleostei; ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            ResSourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 401.
```

```
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3733397"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type "26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']];
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab, ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT       112 a    113 c    103 g     74 t
ORIGIN

  Query Match             6.5%;  Score 93;  DB 12;  Length 402;
  Best Local Similarity   71.1%;  Pred. No. 3.7e-15;
  Matches  123;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy       1163 TTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAACCTCCTGATGC 1222
             || ||| |  ||  | ||||| |||| ||| ||||| || || ||| | || || ||||
Db          1 TTTCCAGAACGGGTCTGTCCTCACAGCGGCAGTCCATCCGGCTCGCTCATCTGCTAATGC 60

Qy       1223 TCCTGTCCCACGTCAGGCATGCGAGTAACAAGGGCATGGAACATCTGCTCAACATCAACT 1282
             | || || ||| || | ||    || ||||| ||||| || |||||   |||||||||
Db         61 TGCTCTCACACATCCGACACCTCAGCAACAAAGGCATCGAGCATCTGTCAAACATGAAGA 120

Qy       1283 GCAAAAATGTGGTCCCAGTGTATGACCTGCTCCTGGAGATGCTGAATGCCCAC 1335
              | ||||| ||||| |   |||| || || ||||| |||||||||| | ||  ||
Db        121 GGAAAAACGTGGTGCTGCTGTACGATCTTCTGCTAGAGATGCTGGACGCAAAC 173


RESULT 33
BM957249
LOCUS       BM957249                 669 bp    mRNA    linear   EST 18-MAR-2002
DEFINITION  ty75h10.y1 Suqano SJD adult male Danio rerio cDNA clone 5777131 5'
            similar to TR:Q90272 Q90272 RETINOIC ACID RECEPTOR ALPHA 2.B. ;
            mRNA sequence.
ACCESSION   BM957249
VERSION     BM957249.1  GI:19540679
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopteryqii; Neopteryqii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: ty75h10.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Suqano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
            High quality sequence stop: 526.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="5777131"
                     /clone_lib="Suqano SJD adult male"
                     /sex="male"
                     /tissue_type="whole body"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACCATGTG);
                     Site_2: DraIII (CACTGTGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed and donated by Dr. Sumio
                     Suqano (University of Tokyo Institute of Medical Science).
                     Custom primers for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA. "
BASE COUNT       182 a    174 c    183 g    130 t
ORIGIN

  Query Match             6.4%;  Score 91.8;  DB 14;  Length 669;
  Best Local Similarity   60.7%;  Pred. No. 1.1e-14;
  Matches  150;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        242 CCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCCCTGTCTGCAGCG 301
              |||||||| | |    ||| | ||    |||       |  |||   |||||||   |
Db        220 CCAGCCCTCCTTCACCGCCACCTCCACCACGGGTCTACAAGCCCTGTTTTGTCTGCCAGG 279

Qy        302 ATTACGCATCGGGATATCACTATGGAGTCTGGTCCGTGTGAAGGATGTAAGGCCTTTTTIA 361
              | | | ||||| || ||||||||||||| | | |||||||| |||||||  |||||
Db        280 ACAAGTCTTCGGGCTACCACTATGGAGTCAGCGCATGTGAAGGGTGTAAGGGTTTTTICC 339

Qy        362 AAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCG 421
              ||||||||||||      | |  | || | || |         ||    |||| |||
Db        340 GAAGAAGCATTCAGAAGAACATGGTGTACACCTGCCACAGAGAGAAGAGCTGTATCATCA 399

Qy        422 ATAAAAACGGGGGCAAGAGCTGCCAGGCTGCGACTTCGGAAGTGTTACGAAGTGGGAA 481
              | ||   |  || ||  |||||||  |||||||| | ||||||    |||||||| |
Db        400 ACAAGGTCACCCGAAACCGCTGCCAGTACTGCCGACTGCAGAAGTGCCTGGAAGTGGGCA 459

Qy        482 TGGTGAA 488
              ||  ||
Db        460 TGTCIAA 466


RESULT 34
BB711891
LOCUS       BB711891                 434 bp    mRNA    linear   EST 11-OCT 2001
DEFINITION  BB711891 RIKEN full length enriched, 2 cells cDNA Mus musculus cDNA
            clone B020040L17 3', mRNA sequence.
ACCESSION   BB711891
VERSION     BB711891.1  GI:16065060
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Konda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            Waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumra
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384 format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer based methods for the mouse full-length cDNA
            encyclopedia: real time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES             Location/Qualifiers
     source          1..434
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="B0200301L17"
                     /clone_lib "RIKEN full length enriched, 2 cells egg"
                     /tissue_type "egg"
                     /dev_stage "2 cells"
                     /lab_host "DH10B"
                     /note "Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN, Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC 1."
BASE COUNT       116 a      99 c     107 g     112 t
ORIGIN

  Query Match             6.4%;  Score 91.2;  DB 10;  Length 434;
  Best Local Similarity   66.0%;  Pred. No. 1.3e-14;
  Matches  142;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGGAAG 343
              ||||| |||||| |||| | || ||||| || |||| || | |||||| |
Db        219 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGGTGCTGACGTGTGGAA 278

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| | ||| |||||||||    | |||||| | |||| |||| | | |
Db        279 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACTTTGTGCTGGAA 338

Qy        404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | || | || ||||||| || | || | ||| | || |||| ||| |
Db        339 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 398

Qy        464 AGTGTTACGAAGTGGGAATG 483
              | |||  ||| ||||||
Db        399 AATGTCTTCAAGCTGGAATG 418


RESULT 35
BG687711
LOCUS       BG687711                 521 bp    mRNA    linear   EST 02-MAY 2001
DEFINITION  344912 BARC 5B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BG687711
VERSION     BG687711.1  GI:13929512
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria, Cetartiodactyla, Ruminantia, Pecora, Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and
            Wells,K.D.
  TITLE     Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@anri.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 33  row: E  column: 5
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..521
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib "BARC 5B0V"
                     /tissue_type "pooled"
                     /lab_host "DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."
BASE COUNT       162 a      111 c      114 g      134 t
ORIGIN

  Query Match             6.4%;  Score 91.2;  DB 12;  Length 521;
  Best Local Similarity   66.0%;  Pred. No. 1.4e-14;
  Matches  142;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGGAAG 343
              |||||  ||||||  ||| | || || |||| ||| ||||| |||| | | ||||
Db        225 TCTGCCTAGTCTGCTCTGATGAAGCTTCAGGATGTCATTATGGGGTCTTGACTTGTGGAT 284

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| |  || || ||||||    | |||||| |  |||| |||| | | |
Db        285 GCTGTAAAGTATTCTTCAAAAGAGCAGTGGAAGGACAGCACAATTATCTTTGTGCTGGAA 344
```

```
Qy          404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
                ||||  |  ||||  ||  |||||||  ||  |  || |  |||||  ||  |||||    |  |
Db          445 GAAATGATTGTATCATTGATAAAATTCGAAGAAAAAACTGCCCAGCATGCCGCTATAGAA 504

Qy          464 AGTGTTACGAAGTGGGAATG 483
                |  ||     |||  ||||||
Db          505 AATGTCTTCAAGCTGGAATG 524


RESULT 36
BB196002
LOCUS       BB196002                 698 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB196002 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330092F23 3' similar to X13358 Mouse mRNA for
            glucocorticoid receptor form A, mRNA sequence.
ACCESSION   BB196002
VERSION     BB196002.2  GI:15409595
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Taqami,M., Taqawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,I.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jun 30, 2000 this sequence version replaced gi:8860955.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro cho, Tsurumi ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             Waqi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system  384 format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, 172-186 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..698
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A330092E23"
                     /clone_lib="RIKEN full-length enriched, adult male spinal
                     cord"
                     /sex="male"
                     /tissue_type="spinal cord"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot   10.0 and subtraction to Rot   450.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATTCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT      193 a    149 c    173 g    183 t
ORIGIN

  Query Match             6.4%;  Score 91.2;  DB 10;  Length 698;
  Best Local Similarity   66.0%;  Pred. No. 1.6e-14;
  Matches  132;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 343
                ||||||   ||  |||  ||||  | ||  |||||||  ||  |||||  ||    |  |||||| |
Db          485 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGTGCTGACGTGTGGAA 544

Qy          344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
                |  ||||| |  |||  |||||||||     |  |||||||  |  ||||  |||||  | |  |
Db          545 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 604

Qy          404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
                |||| |  ||  |   ||  |||||||  ||  |  || |  |||  |   ||  |||||    |||  |
Db          605 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 664

Qy          464 AGTGTTACGAAGTGGGAATG 483
                |  |||     |||  ||||||
Db          665 AATGTCTTCAAGCTGGAATG 684


RESULT 37
BC016462
LOCUS       BC016462                 1638 bp    mRNA    linear   HTC 07-AUG-2002
DEFINITION  Mus musculus, Similar to nuclear receptor subfamily 3, group C,
            member 1, clone IMAGE:4036433, mRNA.
ACCESSION   BC016462
VERSION     BC016462.1  GI:16741246
KEYWORDS    HTC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1638)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-rmail.nih.gov
```

```
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amq@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 24 Row: c Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6680102
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..1638
                     /organism "Mus musculus"
                     /db_xref="taxon:10090"
                     /map="5 73.1 cM"
                     /clone="IMAGE:4036433"
                     /tissue_type "Mammary tumor, WAP-TGF alpha model, 7 months
                     old, gross tissue."
                     /clone_lib "NCI_CGAP_Mam5"
                     /lab_host "DH10B"
                     /note "Vector: pCMV SPORT6"
BASE COUNT      446 a     396 c     443 g     353 t
ORIGIN

  Query Match             6.4%;  Score 91.2;  DB 11;  Length 1638;
  Best Local Similarity   66.0%;  Pred. No. 2.6e-14;
  Matches  142;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAG 343
              ||||||   || |||  |||| | || |||||||  || ||||| ||   | ||||||  |
Db       1403 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGGTGCTGACGTGTGCAA 1462

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| | ||| |||||||||    | |||||||  | ||||  ||||| | |  |
Db       1463 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGACAGCACAATTACCTTTGTGCTGGAA 1522

Qy        404 CAAATCAGTGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | || | || |||||||  ||  | || | ||| |  || |||||   ||| |
Db       1523 GAAATGATTGCATCATGATAAAATTCGAACAAAAAACTGTCCAGCATGCCGCTATCGAA 1582

Qy        464 AGTGTTACGAAGTGGGAATG 483
              | |||    |||  ||||||
Db       1583 AATGTCTTCAAGCTGGAATG 1602


RESULT 38
BG010872
LOCUS       BG010872                399 bp    mRNA    linear   EST 24 JAN 2001
DEFINITION  RC5-GN0194-061200-024-F11 GN0194 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG010872
VERSION     BG010872.1  GI:12458558
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 4491-4496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/qethtml2.pl?I1 RC5&t2=RC5-GN0194
            0&1200-024-F11&13-2000-12-06&t4-1)
            Seq primer: puc 18 forward
            High quality sequence stop: 70.
FEATURES             Location/Qualifiers
     source          1..399
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib="GN0194"
                     /dev_stage "Adult"
                     /note "organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716    Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      107 a      86 c     111 g      95 t
ORIGIN

  Query Match             6.3%;  Score 91;  DB 12;  Length 399;
  Best Local Similarity   63.5%;  Pred. No. 1.4e-14;
  Matches  139;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        292 CTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAG 351
              || ||  | ||   ||| |  || || ||||||||  ||    || |||||||  || |||
Db        138 GTGTGTGGTGACATCGCTTTTGGGTACCACTATGGGGTAGCATCATGTGAAGCCTGCAAG 197

Qy        352 GCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAG 411
              || || ||  || || |  ||||||||  ||  || || |  ||  || || || ||| |
Db        198 GCATTCTTCAAGAGGACAATTCAAGGCAATATAGAATACAGCTGCCCTGCCACGAATGAA 257

Qy        412 TGTACAATCGATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTAC 471
              ||  ||||   ||   | | || ||  |||||||||  |||||  |  |||||||
Db        258 TGTGAAATCACAAAGCGCAGACGTAAATCCTGCCAGGCTTGCCGCTTCATGAAGTGTTTA 317

Qy        472 GAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGA 510
              |||||||  ||| ||||    ||    ||    || |||
Db        318 AAAGTGGGCATGCTGAAAGAAGGGGTGCGTCTTGACAGA 356


RESULT 39
BG746447
LOCUS       BG746447                592 bp    mRNA    linear   EST 15 MAY 2001
DEFINITION  602704728F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4856886 5',
            mRNA sequence.
ACCESSION   BG746447
VERSION     BG746447.1  GI:14057100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Bonq/Rubin Laboratory
```

```
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1708  row: h  column: 07
            High quality sequence stop: 587.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4856886"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT       98 a    182 c    197 g    115 t
ORIGIN

  Query Match             6.3%;  Score 90.6;  DB 12;  Length 592;
  Best Local Similarity   48.5%;  Pred. No. 2.2e-14;
  Matches  278;  Conservative    0;  Mismatches  294;  Indels    1;  Gaps    1;

Qy        285 CTGCCGCTGTCTGCAGCCATTACGCATCGGGATATCACTATGGAGTCTGCTCGTCTCAAGCC 344
              ||||   |||||  | ||     || || || || ||||||| ||     || |||||
Db          2 CTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTATGGTCGGGATACCTGTCAAGGC 61

Qy        345 ATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTAC 404
              || || ||||| || || || | ||| || |||   |  || || |  ||||| ||  |
Db         62 CTGCAAAGCCTTCTTCAAGAGGACCATCCAGGGAGC-ATCGAGTACAGCTGTCCGGCCTC 120

Qy        405 AAATCAGTGTACAATCGATAAAAACCGCCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAA 464
              ||  |||||   |||   ||    | ||||||  ||||||||||||||||  |   ||
Db        121 CAACGAGTGTGAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTCACCAA 180

Qy        465 GTGTTACGAAGTGGGAATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCT 524
              |||       ||||| ||| | |||   ||    ||    ||  |     ||    | |
Db        181 GTGCCTGCGGGTGGGCATGCTCAAGGAGGAGTGCGCCTGGACCGCCTCCCGCCTCCCCC 240

Qy        525 TGTGCGGAGACAGAGAAGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAG 584
              |  |    || |  |  | || | | | |  |        |        ||
Db        241 GCAGAAGTACAAGCGGCGGCCGGAGGTGGACCCACTGCCCTTCCCGGGCCCCTTCCCTGC 300

Qy        585 TGGCGGCCACGCGCCCCGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCT 644
              |||   ||  ||   |   |   |  | |  |  | ||||      |  | | |
Db        301 TGGGCCCCTGGCAGTCGCTGGAGGCCCCCGGAAGACAGCAGCCCCAGTGAATGCACTGGT 360

Qy        645 AGTGCTCACCCTCCTGGAGGCTGAGCCGCCCCATGTGCTGATCAGCGCGCCCAGTGCGCC 704
              |     ||  |||  |       |  |        ||      ||||   |  ||
Db        361 GTCTCATCTGCTGGTGGTGAGCCTGAGAAGCTCTATGCCATGCCTGACCCCGCAGGCCC 420

Qy        705 CTTCACCGAGGCCTCCATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACA 764
                  |  | |     ||    ||||     |  |    |||   ||| | ||
Db        421 TGATGGGCACCTCCCAGCCGTGGCTACCCTCTGTGACCTCTTTGACGGAGAGATTGTGGT 480

Qy        765 CATGATCAGCTGGGCCAAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGT 824
              ||  ||||||||||||||  || || ||||      |||    ||||  |||||  |
Db        481 CACCATCAGCTGGGCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAGAT 540

Qy        825 GCGGCTCTTGGAGAGCTGTTGGATGGAGGTGTT 857
              |    |  || ||||   ||||||||||||| |
Db        541 GTCAGTACTGCAGAGCGTGTGGATGGAGGTGCT 573


RESULT 40
AW113877/c
```

```
LOCUS       AW113877                 413 bp    mRNA    linear   EST 31-JAN 2000
DEFINITION  MC809 mouse liver, vehicle control Mus musculus cDNA clone MC809 3'
            similar to X13458 Mouse mRNA for glucocorticoid receptor form A,
            mRNA sequence.
ACCESSION   AW113877
VERSION     AW113877.1  GI:6826590
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Thomas,R.S., Rank,D.R., Penn,S.G., Zastrow,G.M., Jovanovich,S. and
            Bradfield,C.A.
  TITLE     McArdle Laboratory/Molecular Dynamics Dioxin EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Bradfield CA
            McArdle Laboratory for Cancer Research
            University of Wisconsin
            1400 University Ave., Madison, WI 53706, USA
            Tel: 608 262 2024
            Fax: 608 262 2824
            Email: bradfield@oncology.wisc.edu
            This clone was sequenced as part of a project to develop a database
            on gene expression changes following exposure to various
            environmental toxicants. The database can be accessed at
            http://mcardle.oncology.wisc.edu/bradfield.  Treatment: dioxane
            vehicle (400 ul/kg). Animals sacrified: 48 hrs post-injection.
            Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="MC809"
                     /clone_lib="mouse liver, vehicle control"
                     /sex="male"
                     /lab_host="JM109"
                     /note="Organ: liver; Vector: pGEM11Zf (Promega); Site_1:
                     NotI; Site_2: EcoRI; First strand cDNA was primed with a
                     NotI-polyT primer
                     [5'-AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT-3'].
                     Double-stranded cDNA was ligated with EcoRI adapters
                     (Pharmacia), digested with NotI, and ligated into the
                     EcoRI/NotI sites of the pGEM11Zf(+) vector.  The library was
                     NOT normalized."
BASE COUNT      103 a    105 c     98 g    103 t      4 others
ORIGIN

  Query Match             6.3%;  Score 90.2;  DB 10;  Length 413;
  Best Local Similarity   65.5%;  Pred. No. 2.4e-14;
  Matches  131;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        284 TCTGCGCTGTCTGCAGCGATTACGCATCGGGATATCACTATGGAGTCTGGTCGTCTCAAG 343
              |||||   || |||  |||| | || |||||||  || ||||| ||    | ||||||  |
Db        214 TCTGCCTGGTGTGCTCCGATGAAGCTTCGGGATGCCATTATGGGCTCTCGACGTGTGGAA 155

Qy        344 GATGTAAGGCCTTTTTTAAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTA 403
              | ||||| | ||| |||||||||     | |||| ||  |  ||||  |||| | |  |
Db        154 GCTGTAAAGTCTTCTTTAAAAGAGCAGTGGAAGGNCAGCACAATTACCTTTGTGCTGGAA 95

Qy        404 CAAATCAGTGTACAATCGATAAAAACCGCCGCAAGAGCTGCCAGGCCTGCCGACTTCGGA 463
              |||| | || |  || |||||||  ||  | || | ||| |  || |||||   ||| |
Db         94 GAAATGATTGCATCATTGATAAAATTCGAAGAAAAAACTGTCCAGCATGCCGCTATCGAA 35

Qy        464 AGTGTTACGAAGTGGGAATG 483
              | |||    |||  ||||||
Db         34 AATGTCTTCAAGCTGGAATG 15


Search completed: June 16, 2003, 03:16:05
Job time : 2195.68 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 00:23:52 ; Search time 3414.22 Seconds
(without alignments)
10988.581 Million cell updates/sec

Title: US-08-826-361A-2
Perfect score: 1251
Sequence: 1 ATGAATTACAGCATTCCCAG ACGTCAGGCATGCGAGGTGA 1251

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : GenEmbl:*
1: qb_ba:*
2: qb_htg:*
3: qb_in:*
4: qb_om:*
5: qb_ov:*
6: qb_pat:*
7: qb_ph:*
8: qb_pl:*
9: qb_pr:*
10: qb_ro:*
11: qb_sts:*
12: qb_sy:*
13: qb_un:*
14: qb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1251	100.0	1251	6	A65105	A65105 Sequence 2
2	1251	100.0	1251	6	AX323056	AX323056 Sequence
3	1249	99.8	1257	6	A65123	A65123 Sequence 20
4	1249	99.8	1257	6	AX323074	AX323074 Sequence
5	1248	99.8	2041	9	AF051428	AF051428 Homo sapi
6	1247.4	99.7	2745	9	AF060555	AF060555 Homo sapi
7	1247	99.7	1434	6	A65104	A65104 Sequence 1
8	1247	99.7	1434	6	AX323055	AX323055 Sequence
9	1247	99.7	1460	6	A61586	A61586 Sequence 4
10	1247	99.7	1460	6	AR075921	AR075921 Sequence
11	1247	99.7	1460	6	AX022458	AX022458 Sequence
12	1247	99.7	1560	9	HSRNAERB	X99101 H.sapiens m
13	1247	99.7	1647	6	AX029400	AX029400 Sequence
14	1247	99.7	1740	9	AB006590	AB006590 Homo sapi
15	1247	99.7	1898	6	A65127	A65127 Sequence 24
16	1247	99.7	1898	6	AX323078	AX323078 Sequence
17	1247	99.7	2011	6	AX234658	AX234658 Sequence
18	1247	99.7	2011	9	AF051427	AF051427 Homo sapi
19	1246.4	99.6	3593	9	AB006589	AB006589 Homo sapi
20	1185.6	94.8	1640	9	AF393815	AF393815 Macaca ar
21	1126.4	90.0	1566	9	AF393816	AF393816 Callithri
22	1125.4	90.0	1651	9	CJERB	Y09472 C.jacchus m
23	995	79.5	1584	4	AF177936	AF177936 Ovis arie
24	987	78.9	1611	4	AF110402	AF110402 Bos tauru
25	985.4	78.8	1820	4	BTA18017	Y18017 Bos taurus
26	963	77.0	1581	4	AF164957	AF164957 Sus scrof
27	963	77.0	1581	4	AF267736	AF267736 Sus scrof
28	960	76.7	1427	9	AF124790	AF124790 Homo sapi
29	960	76.7	2153	9	BC024181	BC024181 Homo sapi
30	940.2	75.2	1621	10	AF042058	AF042058 Rattus no
31	940.2	75.2	1650	10	RNAJ2602	AJ002602 Rattus no
32	937	74.9	2555	10	RNU57439	U57439 Rattus norv
33	937	74.9	2568	6	A61583	A61583 Sequence 1
34	937	74.9	2568	6	AR075920	AR075920 Sequence
35	937	74.9	2568	6	AX022455	AX022455 Sequence
36	929	74.3	2152	10	MMU81451	U81451 Mus musculu
37	922.6	73.7	1458	6	A61588	A61588 Sequence 6
38	922.6	73.7	1458	6	AR075922	AR075922 Sequence
39	922.6	73.7	1458	6	AX022460	AX022460 Sequence
40	922.6	73.7	1458	10	MMAJ220	AJ000220 Mus museu
41	876.2	70.0	1512	10	RNAJ2603	AJ002603 Rattus no
42	876.2	70.0	1675	10	AF042059	AF042059 Rattus no
43	873	69.8	1512	10	AB012721	AB012721 Rattus no
44	865	69.1	1512	10	AY054413	AY054413 Mus museu
45	781	62.4	1215	9	AF074599	AF074599 Homo sapi
46	735.8	58.8	3163	5	AF113513	AF113513 Sturnus v
47	718.6	57.4	1504	10	AF042060	AF042060 Rattus no
48	674.2	53.9	3539	5	AB036415	AB036415 Gallus ga
49	671.4	53.7	840	9	AF119229	AF119229 Macaca mu
50	665.6	53.2	1419	5	AF045149	AF045149 Coturnix
51	654.6	52.3	1558	10	AF042061	AF042061 Rattus no
52	616.6	49.3	751	4	CFA313196	AJ313196 Canis fam
53	554.8	44.3	1795	5	AF147746	AF147746 Squalus a
54	439.8	35.2	3061	5	AB003356	AB003356 Anguilla
55	426.2	34.1	716	4	AF257109	AF257109 Ovis arie
56	425.4	34.0	1937	5	DRE414567	AJ414567 Danio rer
57	423.6	33.9	2018	5	AF349414	AF349414 Danio rer
58	413.2	33.0	2850	5	AB070630	AB070630 Paralicht
59	404.8	32.4	2283	5	AF061269	AF061269 Carassius
60	403.2	32.2	1909	5	AB083064	AB083064 Cyprinus
61	402	32.1	2276	5	AB070901	AB070901 Oryzias l
62	394.4	31.5	4306	5	TNU75605	U75605 Tilapia nil
63	394.2	31.5	2508	5	OMY289883	AJ289883 Oncorhync
64	392.4	31.4	2180	5	AF136980	AF136980 Sparus au
65	386.2	30.9	1772	5	AF298182	AF298182 Micropogo

66	375.2	30.0	3804	5	AF185568	AF185568 Ictalurus
67	372.8	29.8	2650	5	AF177465 Carassius	
68	369.4	29.5	2462	5	AF298183 Micropog	
69	361.2	28.9	1925	5	DRE414566	AJ414566 Danio rer
70	339.6	27.1	626	10	AB076607	AB076607 Mesocrice
71	338.8	27.1	2764	6	AX376687	AX376687 Sequence
72	338.8	27.1	2770	5	ORZMER	D28954 Oryzias sp.
73	336.4	26.9	2792	5	ZFIER	L79911 Taeniopygia
74	336.2	26.9	2390	5	AF298183	AF298183 Micropogo
75	334.2	26.7	2821	5	AB037185	AB037185 Danio rer
76	333.2	26.6	2085	5	GGERR	X03805 Chicken mRN
77	332.4	26.6	1728	6	E39777	E39777 Estrogen re
78	332.4	26.6	1863	6	E39778	E39778 Estrogen re
79	331	26.5	3439	5	AF349412	AF349412 Danio rer
80	328.8	26.3	2552	5	DRE275911	AJ275911 Danio rer
81	327.4	26.2	5078	5	TNU75604	U75604 Tilapia nil
82	324.6	25.9	2423	5	AF326201	AF326201 Halichoer
83	324.2	25.8	3416	5	AF136979	AF136979 Sparus au
84	322	25.7	1374	6	AX066402	AX066402 Sequence
85	322	25.7	1788	6	AX411751	AX411751 Sequence
86	322	25.7	4963	6	AR029418	AR029418 Sequence
87	322	25.7	4963	6	115368	115368 Sequence 16
88	322	25.7	4963	6	156760	156760 Sequence 9
89	321.6	25.7	2369	5	SAU6039	AJ006039 Sparus au
90	321.4	25.7	2106	9	HSU47678	U47678 Human 80 kD
91	320.4	25.6	2092	6	AX474709	AX474709 Sequence
92	320.4	25.6	2092	6	108538	108538 Sequence 1
93	320.4	25.6	2092	9	HSMERMCF	M12674 Human estro
94	320.4	25.6	6450	6	AR153585	AR153585 Sequence
95	320.4	25.6	6450	9	HSERR	X03635 Human mRNA
96	320.4	25.6	6610	12	AF061181	AF061181 Mammalian
97	320	25.6	3500	5	AR007453	AR007453 Chrysophr
98	318.8	25.5	5653	4	AY033393	AY033393 Ovis ario
99	317.6	25.4	1966	4	OAESTREC	Z49257 O.aries mRN
100	317.2	25.4	1788	4	SSESTREC	Z37167 S.scrofa mR
101	314	25.1	1764	5	AB055220	AB055220 Caiman cr
102	309.6	24.7	1880	10	RNESTROR	X61098 Rat mRNA fo
103	309.6	24.7	2087	5	AY055725	AY055725 Carassius
104	309.6	24.7	2090	10	S52128	S52128 estrogen re
105	309.6	24.7	2090	10	RNERR	Y09102 Rat mRNA fo
106	307.6	24.6	2796	5	AB070629	AB070629 Paralicht
107	307.6	24.6	3746	4	AF124093	AF124093 Equus cab
108	306.4	24.5	1223	6	AX066401	AX066401 Sequence
109	306.4	24.5	1978	10	AF181077	AF181077 Mesocrice
c 110	302	24.1	127709	2	CNS05TC1	AL352983 Homo sapi
c 111	302	24.1	143340	9	AP215947	AP215947 Homo sapi
112	302	24.1	176257	9	CNS01RBJ	AL161756 Human chr
113	302	24.1	325791	6	AX234657	AX234657 Sequence
114	301.6	24.1	2102	5	CGESTREC	X84743 C.quariepinu
115	298.4	23.9	2078	10	MUSESTRU	M38651 Mouse estro
116	296.8	23.7	6031	5	AF253505	AF253505 Ictalurus
117	296.8	23.7	6756	5	AF061275	AF061275 Ictalurus
c 118	296.8	23.7	8762	5	AF253507	AF253507 Ictalurus
119	296.2	23.7	1746	5	AB055221	AB055221 Chemidoph
120	292	23.4	1955	5	OMY242741	AJ242741 Oncorhync

ALIGNMENTS

RESULT 1
A65105
LOCUS A65105 1251 bp DNA linear PAT 29-MAR-1999
DEFINITION Sequence 2 from Patent EP0798378.
ACCESSION A65105
VERSION A65105.1 GI:4530974
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1251)
AUTHORS Mosselman,S. and Dijkema,R.
TITLE Estrogen receptor
JOURNAL Patent: EP 0798378-A 2 01-OCT-1997;
AKZO NOBEL NV (NL)
COMMENT Other publication: CA 2200423 19970926
Other publication: AU 1652197 19971002.
FEATURES Location/Qualifiers
source 1..1251
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 297 a 324 c 354 g 276 t
ORIGIN

Query Match 100.0%, Score 1251, DB 6, Length 1251;
Best Local Similarity 100.0%; Pred. No. 1.8e-272;
Matches 1251, Conservative 0, Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAACAGCGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAACGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGTGA 1251
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGTGA 1251


RESULT 2
AX323056
LOCUS       AX323056               1251 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 2 from Patent EP1162264.
ACCESSION   AX323056
VERSION     AX323056.1  GI:18093943
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 2 12-DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1..1251
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      297 a    324 c    354 g    276 t
ORIGIN

  Query Match             100.0%;  Score 1251;  DB 6;  Length 1251;
  Best Local Similarity   100.0%;  Prod  No  1 8e-272;
  Matches 1251;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGGCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGGCTGGTCGGCAGACC 60

Qy          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCGCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCGCGCC 600

Qy         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACGGAGGCTCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACGGAGGCTCCC 720

Qy         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGTGA 1251
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGTGA 1251


RESULT 3
A65123
LOCUS       A65123                 1257 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 20 from Patent EP0798378.
ACCESSION   A65123
VERSION     A65123.1  GI:4530988
KEYWORDS    .
```

```
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1257)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378 A 20 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     other publication CA 2209423 19970926
            other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1. .1257
                     /organism "unidentified"
                     /db_xref "taxon:32644"
BASE COUNT       297 a     327 c     355 g     278 t
ORIGIN

  Query Match             99.8%;  Score 1249;  DB 6;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-272;
  Matches 1249;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGACA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGCCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGT 1249
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGT 1249


RESULT 4
AX323074
LOCUS       AX323074                1257 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 20 from Patent EP1162264.
ACCESSION   AX323074
VERSION     AX323074.1  GI:18093957
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 20 12-DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1257
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       297 a     327 c     355 g     278 t
ORIGIN

  Query Match             99.8%;  Score 1249;  DB 6;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-272;
  Matches 1249;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
```

```
                                                                                
Db        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTCCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTCCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 CATAAAAACCGCCCCCAAGACGTCCCAGCCCTGCCGACTTCCGAACTGTTACCAACTCCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGCCCCAAGACGTCCCAGCCCTGCCGACTTCCGAACTGTTACCAACTCCCA 480

Qy        481 ATGGTGAAGTGTGCCTCCCGCAGACAGACAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGCCTCCCGCAGACAGACAGATGTGGGTACCGCCTTGTGCGGAGACAGACA 540

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAACAGAAGTGGCGGCCACGCGCCC 600

Qy        601 CGACTGCGGGACGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGACTGCGGGACGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCGCCCAGTGCCGCTTCACCCAGCCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCGCCCAGTGCCGCTTCACCCAGCCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 780

Qy        781 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGGGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTGCACTTGCTGAACGCCTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACAGGGACCCAGGATGCTGACAGCAGCCGCAAGCTGGCTGCACTTGCTCAACGCCTG 1140

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGACGGCCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGACGGCCATCTCCTCCCAGCAGCAATCCATG 1200

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGT 1249
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGGT 1249

RESULT 5
```

```
AF051428
LOCUS       AF051428                2041 bp    mRNA    linear   PRI 17-MAR-1998
DEFINITION  Homo sapiens estrogen receptor beta2 splice variant mRNA, complete
            cds.
ACCESSION   AF051428
VERSION     AF051428.1  GI:2961558
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2041)
  AUTHORS   Moore,J.T., McKee,D.D., Moore,L.B., Jones,S.A., Su,J.-L.,
            Horne,E.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Cloning, Ligand Binding and Functional Activity of human Estrogen
            Receptor beta Isoforms
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2041)
  AUTHORS   Moore,J.T., McKee,D.D., Moore,L.B., Jones,S.A., Su,J.-L.,
            Horne,E.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .2041
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             419. .1906
                     /codon_start=1
                     /product="estrogen receptor beta2 splice variant"
                     /protein_id="AAC05751.1"
                     /db_xref="GI:2961559"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHPQLSHLYAEP
                     QFSPWCPAPSLEHTLPVNPETLKPKVSGNRGCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRREKCGYRLVPRQRSADEQLHCACKAKPSGGHAPPVREILLDALSPFQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHARAEKASQTLTSFGMKMETLLPEATMEQ"
BASE COUNT      505 a    532 c    544 g    460 t
ORIGIN

  Query Match           99.8%;  Score 1248;  DB 9;  Length 2041;
  Best Local Similarity  100.0%;  Pred. No. 9e-272;
  Matches 1248;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 637

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTCGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 697

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 757

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGC 817

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTCCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTCCAGC 877

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 937
```

```
Qy        961 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 997

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1057

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTCGGTACCGCCTTGTGCGGAGACAGAGA 1117

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 1177

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 1237

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GAGCCTGAGCCCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 1297

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1357

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1417

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTCGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTCGCGCTCAATTGACCACCCCGGCAAG 1477

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1537

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1597

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1657

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAAGCGCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAAGCGCGTG 1717

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1777

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCCAGG 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1778 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGG 1825


RESULT 6
AF060555
LOCUS       AF060555                2745 bp    mRNA    linear   PRI 29-APR-1998
DEFINITION  Homo sapiens estrogen receptor beta 4 isoform mRNA, complete cds.
ACCESSION   AF060555
VERSION     AF060555.1  GI:3091285
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2745)
  AUTHORS   Moore,J.T., McKee,D.D., Slentz-Kesler,K., Moore,L.B., Jones,S.A.,
            Su,J.L., Horne,E.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .2745
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             471. .2012
                     /note="splice variant"
                     /codon_start=1
                     /product="estrogen receptor beta 4 isoform"
                     /protein_id="AAC15234.1"
                     /db_xref="GI:3091286"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVINLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRREPCGYRLVPRQPSADEQLHCAGKAKRSGCHAPPVPFETIDATSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHEEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASSLSLSWFLFMLFEASCHGVPQTPGGAHMSVSPSRSFEAC
                     QQPRE"
BASE COUNT       741 a    665 c    690 g    649 t
ORIGIN


  Query Match           99.7%;  Score 1247.4;  DB 9;  Length 2745;
  Best Local Similarity   99.9%;  Pred. No. 1.3e-271;
  Matches 1248;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 689

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 749

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 809

Qy        181 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 869

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 929

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTCGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GATTACGCATCGGGATATCACTATGGAGTCTCGTCGTGTGAAGGATGTAAGGCCTTTTTT 989

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 1049

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1109

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1169

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 1229

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
```

```
Db       1230 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 1289

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACCGAGGCCTCC 1349

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATCATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATCATCAGCTGGGCC 1409

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1469

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1470 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1529

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1589

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1649

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1709

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCCTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCCTG 1769

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCAAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCAAC 1829

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCCAGCAATCCATG 1254
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1830 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGCT 1878


RESULT 7
A65104
LOCUS       A65104                  1434 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 1 from Patent EP0798378
ACCESSION   A65104
VERSION     A65104.1  GI:4530973
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1434)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 1 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     Other publication CA 2200423 19970926.
            Other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1. .1434
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      347 a    373 c    406 g    308 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 6;  Length 1434;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGGGCTCCATCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGGGCTCCATCCC 120

Qy        121 CAGTTATCACATCTGTATGCGGAAGCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGTTATCACATCTGTATGCGGAAGCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy        181 GAACACACCTTACCTGTAAACACAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACACACCTTACCTGTAAACACAGAGACACTGAAAAGGAAGGTTAGTGGCAACCGTTGC 240

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540

Qy        541 AGTGCCGACGAGCAGCTGCACGTGTGCGCCAAGGCCAAGAGAAGCGGCGGCACGGGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCCGACGAGCAGCTGCACGTGTGCGCCAAGGCCAAGAGAAGCGGCGGCACGGGCCC 600

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCGGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATCATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATCATCAGCTGGGCC 780

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCCTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               |||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 8
AX424055
LOCUS       AX424055                1434 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 1 from Patent EP1162264.
ACCESSION   AX424055
VERSION     AX424055.1  GI:18093942
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264 A 1 12-DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1..1434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       447 a    473 c    406 g    308 t
ORIGIN

  Query Match           99.7%;  Score 1247;  DB 6;  Length 1434;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60

Qy          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120

Qy         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180

Qy         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240

Qy         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300

Qy         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360

Qy         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTCTCCAGCTACAAATCAGTGTACAATC 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTCTCCAGCTACAAATCAGTGTACAATC 420

Qy         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480

Qy         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGCGTACCGCCTTGTGCGGAGACAGAGA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGCGTACCGCCTTGTGCGGAGACAGAGA 540

Qy         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600

Qy         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660

Qy         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720

Qy         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780

Qy         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840

Qy         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900

Qy         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960

Qy         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               |||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247


RESULT 9
A61586
LOCUS       A61586                  1460 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 4 from Patent WO9709348.
ACCESSION   A61586
VERSION     A61586.1  GI:3715917
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1460)
  AUTHORS   Kuiper,G.G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348 A 4 13-MAR-1997;
            KAROBIO AB (SE)
FEATURES             Location/Qualifiers
     source          1..1460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       452 a    479 c    411 g    318 t
ORIGIN
```

```
  Query Match             99.7%;  Score 1247;  DB 6;  Length 1460;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0,  Mismatches    0,  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 86

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 146

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 206

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 266

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 326

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 386

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 446

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 506

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGACAGATCTCGGTACCGCCTTGTGCGGAGACACACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGTGAAGTGTGGCTCCCGGAGAGACAGATCTCGGTACCGCCTTGTGCGGAGACACACA 566

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCCCCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACCCCCCC 626

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCACCTACTGCTCACCCTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTACTGCTCACCCTCCTC 686

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 746

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 806

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCCGGCTCTTGGAGAGC 866

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGGGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGGGCTCAATTGACCACCCCGGCAAG 926

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 986

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1046

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1106

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1166

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATC 1226

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1273


RESULT 10
AR075921
LOCUS       AR075921               1460 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 4 from patent US 5958710.
ACCESSION   AR075921
VERSION     AR075921.1  GI:10002667
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1460)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J.-A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710-A 4 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1..1460
                     /organism "unknown"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 6;  Length 1460;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0,  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 86

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 146

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 206

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 266

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 326

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 386

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 446

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 506

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGACAGATCTCGGTACCGCCTTGTGCGGAGACACACA 540
```

```
Db        507 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 566

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
Db        567 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 626

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
Db        627 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 686

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
Db        687 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 746

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
Db        747 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 806

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
Db        807 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 866

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
Db        867 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 926

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
Db        927 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 986

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
Db        987 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1046

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
Db       1047 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1106

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
Db       1107 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1166

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
Db       1167 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1226

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
Db       1227 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1274


RESULT 11
AX022458
LOCUS       AX022458               1460 bp    DNA     linear   PAT 07 SEP 2000
DEFINITION  Sequence 4 from Patent EP0935000.
ACCESSION   AX022458
VERSION     AX022458.1  GI:10046069
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FEATURES             Location/Qualifiers
     source          1. .1460
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
BASE COUNT      352 a    379 c    411 g    318 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 6;  Length 1460;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
Db          27 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 86

Qy          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
Db          87 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 146

Qy         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
Db         147 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 206

Qy         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
Db         207 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 266

Qy         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
Db         267 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 326

Qy         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
Db         327 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 386

Qy         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
Db         387 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 446

Qy         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
Db         447 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 506

Qy         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
Db         507 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 566

Qy         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
Db         567 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 626

Qy         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
Db         627 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 686

Qy         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
Db         687 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 746

Qy         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
Db         747 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 806

Qy         781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
Db         807 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 866

Qy         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
Db         867 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 926

Qy         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
Db         927 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 986

Qy         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
Db         987 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1046

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
Db        1047 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1106
```

```
Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1166

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1226

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1273


RESULT 12
HSRNAERB
LOCUS       HSRNAERB                1560 bp    mRNA    linear   PRI 05-SEP-1996
DEFINITION  H.sapiens mRNA for estrogen receptor
ACCESSION   X99101
VERSION     X99101.1  GI:1518262
KEYWORDS    estrogen receptor.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1560)
  AUTHORS   Mosselman,S., Polman,J. and Dijkema,R.
  TITLE     ER beta: identification and characterization of a novel human
            estrogen receptor
  JOURNAL   FEBS Lett. 392 (1), 49-53 (1996)
  MEDLINE   96354875
   PUBMED   8769313
REFERENCE   2  (bases 1 to 1560)
  AUTHORS   Mosselman,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-1996) S. Mosselman, N V Organon, Biotechnology
            and Biochemistry, PO box 20 Oss, Molenstraat 110 Oss, 5340 BH,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1..1560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
                     /clone_lib="Clontech; HL1010B"
                     /clone_lib="Clontech; 7414-1"
                     /dev_stage="adult"
     CDS             19..1452
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="CAA67555.1"
                     /db_xref="GI:1518263"
                     /db_xref="SPTREMBL:Q92731"
                     /translation="MNYSIPSNVINLEGGPGRQTTSPNVEWPTPGHLSPLVVHRQLSH
                     LYAEPQKSPWCEARSLEHTLPVNRETLKRKVSGNPCASPVLGPQSKRDAHFCAVCSDY
                     ASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVG
                     MVKGGSRKERGGYRLVPRQRSADEQLHCAGKAKRSGSHAPPVKEELTPADSLEPQLVLT
                     LLEAEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVR
                     LLESCWMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRF
                     RELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSG
                     ISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGC
                     KSSITGSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT       367 a    408 c    445 g    340 t
ORIGIN

  Query Match          99.7%;  Score 1247;  DB 9;  Length 1560;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          19 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 78

Qy          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db          79 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 138

Qy         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         139 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 198

Qy         181 GAACACACCTTACCTGTAAACACAGCAGACACTGAAAAGGAAGCTTAGTGGCAACCCTTGC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         199 GAACACACCTTACCTGTAAACAGACAGACACTGAAAAGGAAGCTTAGTGGCAACCCTTGC 258

Qy         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGGCGTGTCTGCAGC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         259 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGGCGTGTCTGCAGC 318

Qy         301 GATTACGGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         319 GATTACGGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 378

Qy         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         379 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 438

Qy         421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         439 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 498

Qy         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         499 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 558

Qy         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         559 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 618

Qy         601 CGAGTGCGGGAGCGTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         619 CGAGTGCGGGAGCGTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 678

Qy         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         679 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 738

Qy         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         739 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 798

Qy         781 AACAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         799 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 858

Qy         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         859 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 918

Qy         901 CTCATCTTTGCTCCAGATCTTGTTCTCGACAGGATGAGGGGAAATGCGTAGAAGGAATT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         919 CTCATCTTTGCTCCAGATCTTGTTCTCGACAGGATGAGGGGAAATGCGTAGAAGGAATT 978

Qy         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         979 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1038

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1039 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1098

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1099 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1158

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1159 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1218
```

```
Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1219 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1265


RESULT 13
AX029400
LOCUS       AX029400               1647 bp    DNA     linear   PAT 16-SEP 2000
DEFINITION  Sequence 2 from Patent WO9912961.
ACCESSION   AX029400
VERSION     AX029400.1  GI:10190195
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1647)
  AUTHORS   Wilkinson,H.
  TITLE     Estrogen receptor
  JOURNAL   Patent: WO 9912961 A 2 18-MAR-1999;
            MERCK & CO INC (US) ; WILKINSON HILARY (US)
FEATURES             Location/Qualifiers
     source          1. .1647
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
BASE COUNT      401 a    438 c    436 g    372 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 6;  Length 1647;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 273

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 333

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 393

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 453

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 513

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 573

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 633

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 693

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCCGGAGACAGAGA 753

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              |||.||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 813

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 873

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 933

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 993

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        994 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1053

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1113

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGATGAGGGGAAATGCGTAGAAGGAATT 1173

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1233

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1293

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1353

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1354 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1413

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1460


RESULT 14
AB006590
LOCUS       AB006590               1740 bp    mRNA    linear   PRI 05-FEB-1999
DEFINITION  Homo sapiens mRNA for estrogen receptor beta, complete cds.
ACCESSION   AB006590
VERSION     AB006590.1  GI:2911151
KEYWORDS    estrogen receptor beta.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogawa,S., Inoue,S., Watanabe,T., Hiroi,H., Orimo,A., Hosoi,T.,
            Ouchi,Y. and Muramatsu,M.
  TITLE     The complete primary structure of human estrogen receptor beta (hER
            beta) and its heterodimerization with ER alpha in vivo and in vitro
  JOURNAL   Biochem. Biophys. Res. Commun. 243 (1), 122-126 (1998)
  MEDLINE   98139878
REFERENCE   2  (bases 1 to 1740)
  AUTHORS   Ogawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1997) Sumito Ogawa, Saitama Medical School,
            Department of 2nd Biochemistry; 38 Morohongo, Moroyama, Iruma gun,
            Saitama 350-0495, Japan (E-mail:suogawa@saitama-med.ac.jp,
            Tel:81-492-76-1490, Fax:81-492-94-9751)
FEATURES             Location/Qualifiers
     source          1. .1740
                     /organism "Homo sapiens"
```

```
                     /db_xref="taxon:9606"
     CDS             99. .1691
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="BAA24953.1"
                     /db_xref="GI:2911152"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPCHLSPLVVHRQLSHLYAEP
                     QKSPWCFAASIPSTLPVNRFTLERKVSGNRCASPVIGPQSKFPDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SRREKCGYRLVRRQRSADEQLHCACSAKRSGGHAPPVRFLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSII
                     GSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT       425 a    463 c    459 g    393 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 9;  Length 1740;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 317

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 377

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 437

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGCAAGGTTAGTGGGAACCGTTGC 497

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGGGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGGGCTGTCTGCAGC 557

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 617

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 677

Qy        421 GATAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 GATAAAAACCGGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 737

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 797

Qy        541 AGTGCCGACGAGCAGCTGCACTGTCCGGGCAAGGCCAAGAGAAGTGGGGGGCACGGGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 AGTGCCGACGAGCAGCTGCACTGTCCGGGCAAGGCCAAGAGAAGTGGGGGGCACGGGCCC 857

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTGCTG 917

Qy        661 GAGGCTGAGCCCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 GAGGCTGAGCCCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 977

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1037

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1097

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1157

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAAIGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATT 1217

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1277

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 CACAAACAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1337

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1397

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1457

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1504


RESULT 15
A65127
LOCUS       A65127                  1898 bp    DNA     linear   PAT 29-MAR-1999
DEFINITION  Sequence 24 from Patent EP0798378.
ACCESSION   A65127
VERSION     A65127.1  GI:4530991
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1898)
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Estrogen receptor
  JOURNAL   Patent: EP 0798378-A 24 01-OCT-1997;
            AKZO NOBEL NV (NL)
COMMENT     Other publication CA 2200423 19970926
            Other publication AU 1652197 19971002.
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism "unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       452 a    506 c    501 g    439 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 6;  Length 1898;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415

Qy        181 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482
```

```
RESULT 16
AX823078
LOCUS       AX823078                1898 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 24 from Patent EP1162264.
ACCESSION   AX823078
VERSION     AX823078.1  GI:18094960
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mosselman,S. and Dijkema,R.
  TITLE     Chimeric hormone receptor
  JOURNAL   Patent: EP 1162264-A 24 12-DEC-2001;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      452 a    506 c    501 g    439 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 6;  Length 1898;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 295

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 355

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 415

Qy        181 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GAACACACCTTACCTGTAAACAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 475

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        476 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 535

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 595

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 655

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 715

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 775

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 835

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 895
```

```
Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 955

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1015

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1075

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1135

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1195

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1255

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1315

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1375

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1435

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1436 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1482


RESULT 17
AX234658
LOCUS       AX234658                2011 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 2 from Patent WO0162793.
ACCESSION   AX234658
VERSION     AX234658.1  GI:15593629
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Kalush,F., Cassel,M.J., Hwang,S.S. and Winn-Deen,E.S.
  TITLE     Estrogen receptor beta variants and methods of detection thereof
  JOURNAL   Patent: WO 0162793-A 2 30-AUG-2001
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1..2011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      470 a    541 c    561 g    439 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 6;  Length 2011;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 637

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 697

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 757

Qy        181 CAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 CAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 817

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 877

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 937

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 997

Qy        421 GATAAAAACCGGCGCCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GATAAAAACCGGCGCCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1057

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1117

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACGCGCCC 1177

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACGCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACGCTCCTG 1237

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 1297

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1357

Qy        781 AACAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1417

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1477

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 1537

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1597

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1657

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1717

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        1718 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1777

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db        1778 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1824


RESULT 18
AF051427
LOCUS       AF051427                2011 bp    mRNA    linear   PRI 14-JUL-1999
DEFINITION  Homo sapiens estrogen receptor beta mRNA, complete cds.
ACCESSION   AF051427
VERSION     AF051427.1  GI:2970563
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2011)
  AUTHORS   Moore,J.T., McKee,D.D., Slentz-Kesler,K., Moore,L.B., Jones,S.A.,
            Horne,E.L., Su,J.L., Kliewer,S.A., Lehmann,J.M. and Willson,T.M.
  TITLE     Cloning and characterization of human estrogen receptor beta
            isoforms
  JOURNAL   Biochem. Biophys. Res. Commun. 247 (1), 75-78 (1998)
  MEDLINE   98300286
   PUBMED   9636657
REFERENCE   2  (bases 1 to 2011)
  AUTHORS   Moore,J.T., McKee,D.D., Moore,L.B., Jones,S.A., Su,J.-L.,
            Slentz-Kesler,K., Horne,E.L., Kliewer,S.A., Lehmann,J.M. and
            Willson,T.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1998) Molecular Sciences, Glaxo Wellcome, 5 Moore
            Drive, RTP, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..2011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     CDS             419..2011
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAC05985.1"
                     /db_xref="GI:2970564"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEAKSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSIT
                     GSECSPAEDSKSKEGSQNPQSQ"
BASE COUNT      470 a    541 c    561 g    439 t
ORIGIN

  Query Match             99.7%;  Score 1247;  DB 9;  Length 2011;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-271;
  Matches 1247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         578 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 637

Qy          61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         638 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 697

Qy         121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGACTCCCTGGTGTGAAGCAAGATCGCTA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         698 CAGTTATCACATCTGTATGCGGAACCTCAAAAGACTCCCTGGTGTGAAGCAAGATCGCTA 757

Qy         181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         758 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 817

Qy         241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         818 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 877

Qy         301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         878 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 937

Qy         361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         938 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 997

Qy         421 CATAAAAACCCCCGCAACAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         998 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1057

Qy         481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1058 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1117

Qy         541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACGCGCCC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1118 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACGCGCCC 1177

Qy         601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1178 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 1237

Qy         661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1238 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 1297

Qy         721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1298 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1357

Qy         781 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1358 AAGAAGATTCCCGGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1417

Qy         841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1418 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1477

Qy         901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1478 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 1537

Qy         961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1538 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1597

Qy        1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1598 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1657

Qy        1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1658 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1717

Qy        1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1718 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1777

Qy        1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db        1778 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1824
```

```
RESULT 19
AB006589
LOCUS       AB006589                3593 bp    mRNA    linear   PRI 28-JUL-1998
DEFINITION  Homo sapiens mRNA for estrogen receptor beta cx, complete cds.
ACCESSION   AB006589
VERSION     AB006589.1  GI:3345671
KEYWORDS    estrogen receptor beta cx.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogawa,S., Inoue,S., Watanabe,T., Orimo,A., Hosoi,T., Ouchi,Y. and
            Muramatsu,M.
  TITLE     Molecular cloning and characterization of human estrogen receptor
            betacx: a potential inhibitor ofestrogen action in human
  JOURNAL   Nucleic Acids Res. 26 (15), 3505-3512 (1998)
  MEDLINE   98337908
REFERENCE   2  (bases 1 to 3593)
  AUTHORS   Ogawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1997) Sumito Ogawa, Saitama Medical School,
            Department of 2nd Biochemistry; 38 Morohongo, Moroyama, Iruma-gun,
            Saitama 350-0495, Japan (E-mail:suogawa@saitama-med.ac.jp,
            Tel:81-492-76-1490, Fax:81-492-94-9751)
FEATURES             Location/Qualifiers
     source          1. .3593
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             1276. .2763
                     /codon_start=1
                     /product="estrogen receptor beta cx"
                     /protein_id="BAA31966.1"
                     /db_xref="GI:3345672"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPCHISPIVVRPQLSHIYAFP
                     QKSPWCEAASLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRFRKCYEVGMVKG
                     SRRERCGYRLVPRQRSADFQLHCAGKAKRSGGHAPPVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVFLSLFDQVRLLESC
                     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHARAEKASQTLTSFGMKMETLLPEATMEQ"
BASE COUNT       971 a    811 c    957 g    854 t
ORIGIN

  Query Match             99.6%;  Score 1246.4;  DB 9;  Length 3593;
  Best Local Similarity   99.9%;  Pred. No. 2.1e-271;
  Matches 1247;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 1494

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 1554

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 1614

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 1674

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 1734

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 1794

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 1854

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 1914

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 1974

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 2034

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 2094

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 2154

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 2214

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 2274

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 TGTTGCATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 2334

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 2394

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 2454

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2455 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 2514

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2515 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 2574

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2575 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 2634

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGG 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2635 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGG 2682


RESULT 20
AF393815
LOCUS       AF393815                1640 bp    mRNA    linear   PRI 05-JUL-2001
DEFINITION  Macaca arctoides estrogen receptor beta 2 (ERBeta2) mRNA, complete
            cds.
ACCESSION   AF393815
VERSION     AF393815.1  GI:14600266
KEYWORDS    .
SOURCE      Macaca arctoides.
  ORGANISM  Macaca arctoides
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1640)
  AUTHORS   Scobie,G.A., Wilson,J.A., Millar,M.R., Macpherson,S. and
            Saunders,P.T.
  TITLE     The estrogen receptor beta variant ERbeta cx/ERbeta2 is expressed
            in a wide range of tissues in both Old and New World primates
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1640)
  AUTHORS   Scobie,G.A. and Saunders,P.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) Medical Research Council, Human
            Reproductive Sciences Unit, 37 Chalmers Street, Edinburgh EH3 9ET,
            UK
FEATURES             Location/Qualifiers
     source          1..1640
                     /organism="Macaca arctoides"
                     /db_xref="taxon:9540"
     gene            1..1640
                     /gene="ERbeta2"
     CDS             38..1537
                     /gene="ERbeta2"
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta 2"
                     /protein_id="AAK71417.1"
                     /db_xref="GI:14600267"
                     /translation="MDIKNSPSSLNSPSSYNCSQSILPLEHGSIYIPSSYVESHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTASPNVLWPTPGHLSPLAVHRQLSHLYAEP
                     QKSPWCEAKSLEHTLPVNRETLKRKVSGNRGTSPVTSPSSKRDAHFCAVCSDEASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGG
                     SRRERGGYRLVRRQRSADEQVHCASKAKRSGSHTPLVRELLLDALSPEQLVLTLLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKL
                     QHKEYLCVKAMILLNSNMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQ
                     QSMRLANLLMLLSHVRHARAEKASHSLTSFGMEDGDTLLPEATDGSVTF"
BASE COUNT      406 a    436 c    413 g    384 t       1 others
ORIGIN

  Query Match             94.8%;  Score 1185.6;  DB 9;  Length 1640;
  Best Local Similarity   96.9%;  Pred. No. 1.2e-257;
  Matches 1209;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 256

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||| ||||||||||||||  |||||||||
Db        257 GCAAGCCCAAATGTGTTGTGGCCAACACCTGGACACCTTTCTCCTTTAGCAGTCCATCGC 316

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              |||||||||||||| ||||| ||||||||||||||||||||||||||||||| |||
Db        317 CAGTTATCACATCTATATGCAGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 376

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        377 GAACACACCTTACCTGTAAACAGAGAGACATTGAAAAGGAAGGTTAGTGGGAACCGTTGC 436

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||| ||||| |||||||||||  |||||||||||||||||||||||||
Db        437 ACCAGCCCTGTTACTAGTCCAAGTTCAAAGAGGGACGCTCACTTCTGCGCTGTCTGCAGC 496

Qy        301 CATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 CATTCGCATCGGGATATCACTATGCAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 556

Qy        361 AAAAGAAGCATTCAAGCACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        557 AAAAGAAGCATTCAAGCCATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 616

Qy        421 CATAAAAACCGCCCCAAGAGCTGCCAGCCTGCCCACTTCGGAAGTGTTACGAAGTGGGA 480
              |||||||| |||||||||||||||||||||||||||||||||||||| |||||||||

Db        617 GATAAAAATCGGCGCAAGAGCTGCCAGCCTGCCGACTTCGGAAGTGTTATGAAGTGGGA 676

Qy        481 ATGGTGAAGTGTGGCTGCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 ATGGTGAAGTGTGGCTGCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 736

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGCGCC 600
              |||||||||||||  ||||||| ||| |||||||||||| |||||| |||| |||||
Db        737 AGTGCCGACGAGCAAGTGCACTGCGCCAGCAAGGCCAAGAGGAGTGGCAGCCACACGCCC 796

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              | ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        797 CTAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTGGTGCTCACCCTCCTG 856

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCTCCC 720
              |||||||||| |||||||| ||||||||||||||| |||||||||| |||||||||
Db        857 GAGGCTGAGCCACCCCATGTACTGATCAGCCGCCCCAGCGCGCCCTTCACTGAGGCTCCC 916

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        917 ATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 976

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1036

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1037 TGTTGGATGGAGGTGTTAATGGTGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1096

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGATGAGGGGAAATGCGTAGAAGGAATT 960
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1097 CTCATCTTGGCTCCAGATCTTGTTCTGGATAGGATGAGGGGAAATGCGTAGAAGGAATT 1156

Qy        961 CTGCAAATCTTTCACATGCTCCTGGCAACTACTTCAAGGTTTCCACAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1157 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1216

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1217 CACAAAGAGTATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAATATGTACCCTCTG 1276

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1277 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCCCACTTGCTGAACGCCGTG 1336

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              ||||| ||||||| ||||||| |||||||||||||||||||||||||||||||||||||
Db       1337 ACCGACGCTTTGGTCTGGGTGATGGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1396

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGG 1248
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1397 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAGG 1444


RESULT 21
AF393816
LOCUS       AF393816                1566 bp    mRNA    linear   PRI 05-JUL-2001
DEFINITION  Callithrix jacchus estrogen receptor beta 2 (ERbeta2) mRNA,
            complete cds.
ACCESSION   AF393816
VERSION     AF393816.1  GI:14600268
KEYWORDS    .
SOURCE      Callithrix jacchus.
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
            Callithrix.
REFERENCE   1  (bases 1 to 1566)
  AUTHORS   Scobie,G.A., Wilson,J.A., Millar,M.R., Macpherson,S. and
            Saunders,P.T.
  TITLE     The estrogen receptor beta variant ERbeta cx/ERbeta2 is expressed
            in a wide range of tissues in both Old and New World primates
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1566)
  AUTHORS   Scobie,G.A. and Saunders,P.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) Medical Research Council, Human
            Reproductive Sciences Unit, 37 Chalmers Street, Edinburgh EH3 9EW,
            UK
FEATURES             Location/Qualifiers
     source          1..1566
                     /organism="Callithrix jacchus"
                     /db_xref="taxon:9483"
     gene            1..1566
                     /gene="ERBeta2"
     CDS             52..1512
                     /gene="ERBeta2"
                     /note="nuclear receptor"
                     /codon_start=1
                     /product="estrogen receptor beta 2"
                     /protein_id="AAK71318.1"
                     /db_xref="GI:14600269"
                     /translation="MDIKNSPSSLNSPSSYNFGQSILPLEHGPIYIPSSYVESHHEYP
                     AMTFYSPAVMNYSIPSSVTNLEEGPGRQITSPNMIWSTPGHLSPLAVHHQLSHLYAEP
                     QKSPWCEARSLEHTLPVSRETLKRKVSGNHCASPVTGPSSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKG
                     SRRERGGYRLVRRQGNAFFQLHCAGKAKRSGGHVPPVRELLLSALSPEQLVLTLLEAE
                     PPHVLISPPSVPFTEASMMMSLTKLADEELVHMISWAKKIPGFVELSLFDQVRLLESC
                     WLEVLMVGLMWRSIDHPGKLIFAPNLLLDRDEGKCVEGILEVFDMLLATTSRFRELKL
                     QHKEYLCVKAMVLLNSQVYPLVTATQDAESSQKLAHLLNAVTDALVWVIAKSGFSSQQ
                     QSVRLANLLMLLSHIRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSI
                     TGSECSPAEDSKSKEIGSSLQDVMQLAQSYSRMNLQDTEYVVLKAIALLNSDAIGIE
                     NKSQIEESYRVTLDHVHKETRHQGSPPRSGDKTDFEKLLSVMPEIRSLVNLTMEKLE
                     VIYVKAKKAKNAVLFDKLLMKVTDLRTINTMHSEKLQSYKVSVEVPDMLSEIIQEQR
                     HLGLSSSPGHLYYQPIH"
BASE COUNT      378 a    410 c    411 g    366 t      1 others
ORIGIN

  Query Match             90.0%;  Score 1126.4;  DB 9;  Length 1566;
  Best Local Similarity   93.9%;  Pred. No. 2.6e-214;
  Matches 1172;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy          1 AtGAATtACAGCATtCCCAGCAATGTCACtAACTTGGAAGGTGGGCCTGGTCGGCAGACC  60
              ||||||||||||||||||| ||| ||||||||||||  ||||||||||||||||||| |
Db        211 ATGAATTACAGCATTCCCAGCAGTGTTACTAACTTGGAAGAGGGGCCTGGTCGGCAGATC 270

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCAGCTTTCTCCTTTAGTGGTCCAtCGC 120
              |||||||||||| || ||||| |||||||||| ||||||||||||||||  ||||||| |
Db        271 ACAAGCCCAAATAtGCTGTGGTCAACACCTGGGCACGTTTCTCCTTTAGCTGTCCATCAC 330

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||| |||
Db        331 CAGTTATCACATCTGTATGCAGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 390

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||| |||||||||||| ||||||||||||||||||||| |||||||||||| |||
Db        391 GAACATACCTTACCTGTAAGCAGAGAGACACTGAAAAGGAAAGTTAGTGGGAACCATTGT 450

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGGCTGTCTGCAGC 300
              |||||||||||||||||||| |||||||||||||||||||||||||| || |||||||||
Db        451 GCCAGCCCTGTTACTGGTCCAAGTTCAAAGAGGGATGCTCACTTCTGGCAGTCTGCAGC 510

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTCTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTCTGTGAAGGATGTAAGGCCTTTTTT 570

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 630

Qy        421 GATAAAAACCGGGGCAAGAGGTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        631 GATAAAAACCGGGGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGCTACGAAGTGGGA 690

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGACACACACA 540
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        691 ATGGTAAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGACACAGGA 750

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGGCC 600
```

```
              | |||||| |||||||||| || ||||||||||||||||||||||| || |||| ||||
Db        751 AATGCCGAGGAGCAGCTGCATTGCGCCGGCAAGGCCAAGAGAAGTGGTGGTCACGTGCC 810

Qy        601 CGAGTGCCGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              |||||||||||||||||||  |||||||||||||||||||||||| ||||||||||||||
Db        811 CGAGTGCCGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTGGTGCTCACCCTCCTG 870

Qy        661 CACCCTCACAGCCCCCACATCGCCGAATCAGCCTTCCAGAAGCTTCTCCAGGTCTTC 720
              ||||| || |||||||||||||||||||||||||| | |||||||| ||||||||
Db        871 GAGGCCGAACCGCCCCATGTGCTGATCAGCCGCCCCAGCGTGCCGTTCACTGAGGCCTCC 930

Qy        721 ATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||
Db        931 ATGATGATGTCCCTGACCAAGCTGGCCGACAGGGAGTTGGTACACATGATCAGCTGGGCC 990

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        991 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 1050

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || ||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1051 TGCTGGTTGGAGGTGTTAATGGTGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1110

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||| ||||  |||||||||||||||||||||||||||||||||||||||
Db       1111 CTCATCTTTGCTCCAAATCTCATTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 1170

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGACTTAAAACTCCAA 1020
              |||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1171 CTGGAAGTCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTGAAACTCCAA 1230

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||| |||||||||||||||||||| |||||||||||||        |||||||
Db       1231 CACAAAGAGTATCTCTGTGTCAAGGCCATCGTCCTGCTCAATTCCCAGTATGACCCTCTG 1290

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCCTG 1140
              ||||||| |||||||||||||| ||||||| |||||||| |||||||||| ||||||
Db       1291 GTCACAGCAACCCAGGATGCTGAGAGCAGCCAGAAGCTGGCCCACTTGCTGAATGCCGTG 1350

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || || |||||||| |||||||| || |||| ||| ||||||||||||||| ||  ||
Db       1351 ACTGACGCTTTGGTCTGGGTGATCGCTAAGAGTGGCTTCTCCTCCCAGCAGCAGTCGGTG 1410

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACATCAGGCATGCGAGG 1248
              |||||||||||||||||||||||||||||||||| ||||||||||||||
Db       1411 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACATCAGGCATGCGAGG 1458


RESULT 22
CJERB
LOCUS       CJERB                   1651 bp    mRNA    linear   PRI 11-NOV-1999
DEFINITION  C.jacchus mRNA for estrogen receptor beta.
ACCESSION   Y09372
VERSION     Y09372.2  GI:6433823
KEYWORDS    DNA binding; estrogen receptor; steroid hormone receptor;
            transactivation.
SOURCE      Callithrix jacchus.
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
            Callithrix.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Gaughan,J.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1651)
  AUTHORS   Gaughan,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1996) J. Gaughan, Medical Research Council,
            Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh, EH3
            9EW, UK
REFERENCE   3  (bases 1 to 1651)
  AUTHORS   Scobie,G.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (03-NOV-1999) G. Scobie., Medical Research Council,
            Centre for Reproductive Biology, 37 Chalmers Street, Edinburgh, EH3
            9EW, UK
COMMENT     on Nov 15, 1999 this sequence version replaced gi:1669661.
            Related entries: X99101 & D57439.
FEATURES             Location/Qualifiers
     source          1..1651
                     /organism="Callithrix jacchus"
                     /db_xref="taxon:9483"
                     /clone="pGEM T vector"
                     /tissue_type="prostate"
                     /tissue_type="ovary"
                     /tissue_type="testis"
     CDS             52..1644
                     /function="steroid hormone receptor"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="CAA70546.2"
                     /db_xref="GI:6438824"
                     /db_xref="SPTREMBL:Q95171"
                     /translation="MDIKNSPSSLNSPSSYNFGQSTLPLEHGPIYIPSSYVESHHEYP
                     AMTFYSPAVMNYSIPSSVTNLEEGPGRQITSPNMLWSTPGHLSPLAVHHQLSHLYAEP
                     QKSPWCEAIGSLEHILPVSKELKKKVSGNHCASPVTGPSSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPAINQCTIDKNRRKSCQACRLKKCYEVGMVKCG
                     SRRERCGYRLVRRQGNAEEQLHCACKKKRSGGHVPRVRELLLSALSPEQLVLTLLEAE
                     PPHVLISRPSVPFTEASMMMSLTKLADEELVHMISWAKKIPGFVELSLLDQVRLLESC
                     WLEVLMVGLMWRSIDHPGKLIFAPNLLLDRDEGKCVEGILEVFDMLLATTSRFRELKL
                     QHKEYLCVKAMVLLNSQYDPLVTATQDAESSQKLAHLLNAVTDALVWVIAKSGFSSQQ
                     QSVRLANLLMLLSHIRHASNKGMEHLLSMKCKNVVPVYDLLLEMMNAHVLRGCKSSIT
                     GSECSPAEDSKSTEQSQNPQSP"
BASE COUNT       397 a    434 c    443 g    377 t
ORIGIN

  Query Match             90.0%;  Score 1125.4;  DB 9;  Length 1651;
  Best Local Similarity   93.9%;  Pred. No. 4.8e-244;
  Matches 1171;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||| ||| ||||||||||||  ||||||||||||||||| |
Db        211 ATGAATTACAGCATTCCCAGCAGTGTTACTAACTTGGAAGAGGGCCTGGTCGGCAGATC 270

Qy         61 ACAAGCCCAAATGTGTTGTGCCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||| ||  |||| ||||||||| |||||||||||  ||||||| |
Db        271 ACAAGCCCAAATATGCTGTGGTCAACACCTGGCCACCTTTCTCCTTTAGCTGTCCATCAC 330

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||-|||||||||||||||| ||||||||||||||||||||||||||||||||||||| |||
Db        331 CAGTTATCACATCTGTATGCAGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCACTA 390

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||| |||||||||||| ||||||||||||||||||| ||||||||||||  |||
Db        391 GAACATACCTTACCTGTAAGCAGAGAGACACTGAAAAGGAAAGTTAGTGGGAACCATTGT 450

Qy        241 GCCAGCCCTGTTACTGGTCCAGTTCAAAGAGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||| || |||||||
Db        451 GCCAGCCCTGTTACTGGTCCAAGTTCAAAGAGGGATGCTCACTTCTGTGCAGTCTGCAGC 510

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GATTACGCATCGGGATATCACTATGGAGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 570

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 630

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        631 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGCTACGAAGTGGGA 690

Qy        481 ATGGTGAAGTGTGGCTCCCCGGAGAGAGAGATGTGGTTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        691 ATGGTAAAGTGTGGCTCCCCGGAGAGAGAGATGTGGTTACCGCCTTGTGCGGAGACAGGGA 750

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              | |||||| |||||||||| || ||||||||||||||||||||||| || |||| ||||
Db        751 AATGCCGAGGAGCAGCTGCATTGCGCCGGCAAGGCCAAGAGAAGTGGTGGTCACGTGCCC 810

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              |||||||||||||||||||  |||||||||||||||||||| ||||||||||||||||
Db        811 CGAGTGCGGGAGCTGCTGCTGAGCGCCCTGAGCCCCGAGCAGCTGGTGCTCACCCTCCTG 870

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||| || |||||||||||||||||||||||||||| | ||||||||| ||||||||
Db        871 GAGGCCGAACCGCCCCCATGTGCTGATCAGCCGCCCCAGCGTGCCCTTCACTGAGGCCTCC 930

Qy        721 ATGATGATCTCCCTCACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        931 ATGATGATGTCCCTGACCAAGCTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 990

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        991 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTGCGGCTCTTGGAGAGC 1050

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || ||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1051 TGCTGGTTGGAGGTGTTAATGGTGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 1110

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              |||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||
Db       1111 CTCATCTTTGCTCCAAATCTCATTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 1170

Qy        961 CTGCAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       1171 CTGCAAGTCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTGAAACTCCAA 1230

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||||| |||||||||||||||||||||| |||||||||||      |||||||
Db       1231 CACAAAGAGTATCTCTGTGTCAAGGCCATGGTCCTGCTCAATTCCCAGTATGACCCTCTG 1290

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||| |||||||||||||| ||||||| |||||||| |||||||||| ||||||
Db       1291 GTCACAGCAACCCAGGATGCTGAGAGCAGCCAGAAGCTGGCCCACTTGCTGAATGCCGTG 1350

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || || |||||||| |||||||| || |||| ||| ||||||||||||||| ||  ||
Db       1351 ACTGACGCTTTGGTCTGGGTGATCGCTAAGAGTGGCTTCTCCTCCCAGCAGCAGTCGGTG 1410

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              |||||||||||||||||||||||||||||||||| ||||||||||||
Db       1411 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACATCAGGCATGCGAG 1457


RESULT 23
AF177936
LOCUS       AF177936                1584 bp    mRNA    linear   MAM 16-JUL-2001
DEFINITION  Ovis aries estrogen receptor beta mRNA, complete cds.
ACCESSION   AF177936
VERSION     AF177936.1  GI:5911382
KEYWORDS    .
SOURCE      Ovis aries.
  ORGANISM  Ovis aries
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 1584)
  AUTHORS   Cardenas,H., Burke,K.A., Bigsby,R.M., Pope,W.F. and Nephew,K.P.
  TITLE     Estrogen receptor beta in the sheep ovary during the estrous cycle
            and early pregnancy
  JOURNAL   Biol. Reprod. 65 (1), 128-134 (2001)
  MEDLINE   21313442
   PUBMED   11420232
REFERENCE   2  (bases 1 to 1584)
  AUTHORS   Cardenas,H., Bigsby,R.M., Pope,W.F. and Nephew,K.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1999) Obstetrics and Gynecology, Indiana
            University, 1001 Walnut St., Indianapolis, IN 46202, USA
```

```
FEATURES             Location/Qualifiers
     source          1. .1584
                     /organism="ovis aries"
                     /db_xref="taxon:9940"
                     /tissue_type="ovary"
     CDS             1. .1584
                     /note="steroid receptor"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAD55772.1"
                     /db_xref="GI:5911383"
                     /translation="MDVKNSPSSLNSPVSYNPGQSTFEFEQHYTLEESYVPESPKEVS
                     AVTFYSPAVMNYSIPNNSEDGPGRQTTSPNVLWPTPGHLSPLAIHCQSSLLYAEPQKS
                     PWCETKSLEHTFEVNKETLEEKASGSPASPVSSPSSKRDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRR
                     ERCGYRIVRRQRNSDPQLHLGSKTKRNGAPMTPVKFLLSALSPPQIVLTLLEAEPPH
                     VLMSRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHK
                     EYLCVKAMILLNSSMYPLVTATQEADSSRKLTHLLNAVTDALVWVIAKSGMSSQQQSM
                     RLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSE
                     RNLAEDSESKEGSQKPQAQ"
BASE COUNT       397 a    420 c    414 g    353 t
ORIGIN

  Query Match             79.5%;  Score 995;  DB 4;  Length 1584;
  Best Local Similarity   87.9%;  Pred. No. 1.5e-214;
  Matches 1085;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy          13 ATTCCCAGCAATGTCACTAACTTGCAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT  72
               | |  ||| | |  || |||||  |||| ||||||||||||| ||||||||||||||||||
Db         163 AATTACAGTATTCCCAATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGCCCAAAT 222

Qy          73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT 132
               || |||||||||||| ||||| ||||| ||||||||||   ||||| ||||||  |||| |
Db         223 GTATTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGTCGTCACTT 282

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTA 192
               |||||||| |||||||||||||||| |||||||| | |||||  ||||||||||||||
Db         283 CTGTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAACCAGATCATTAGAACACACCTTC 342

Qy         193 CCTGTAAACAGAGACACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCCCCAGCCCTGTT 252
               ||||| ||||| |||||| |||||||||| | |||||||| | |||| ||||||||||||
Db         343 CCTGTGAACAGGGAGACATTGAAAAGGAAAGCTAGTGGGAGCAGTTCTGCCAGCCCTGTT 402

Qy         253 ACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCG 312
               || |||  |||||||||||||||| |||||||| |||||||||||||||||| || |||
Db         403 TCTAGTCCGACTTCAAAGAGGGATCCCCACTTCTGTGCTGTCTGCAGCGATTATGCGTCG 462

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAACCATT 372
               ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         463 GGATATCACTATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAACAAGCATT 522

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCCG 432
               |||||||||||||||||||||||||||||||||||||||||||| || ||||| |||||
Db         523 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACCATACATAAGAACCCA 582

Qy         433 CGCAAGAGCTGCCAGGGCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGT 492
               || || ||||| |||||||| ||||| ||||||| || || || ||||||||||||||||
Db         583 CGTAAAAGCTGTCAGGCCTGTCGACTGCGGAAGTGCTATGAGGTCGGAATGGTGAAGTGT 642

Qy         493 GGCTCCCGGAGACAGACAGATGTGGGTACCGCCTTGTGCCGGAGACAGAGAACTCCCCACCAC 552
               |||||||||||||| || |||||||||||| | ||||| || ||||||| | | || |||
Db         643 GGCTCCCGGAGAGAAAGGTGTGGGTACCGCATAGTGCGAAGGCAGAGAAATTCTGATGCAC 702

Qy         553 CAGCTGCACTGTGCCGGCAAGGCCAAGACAAGTGCGCGCCACGCGCCCCCAGTGCCGGAC 612
               ||||||||||||     ||||  ||||||| | || |  | |  | |||||||  ||||
Db         703 CAGCTGCACTGCCTGAGCAAAACCAAGAGGAACGGTGCACCCATGACCCGAGTGAACGAC 762

Qy         613 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 672
               |||||||||  ||||||||||| ||||||| |||| || ||||||||||| |||||||||
Db         763 CTGCTGCTGAGTGCCCTGAGCCCAGAGCAGCTGGTGCTTACTCTCCTGGAGGCCGAGCCG 822

Qy         673 CCCCATGTGCTGATCAGCGCCCCAGTGCCGCTTCACCGAGGCTCCATGATGATGTCC 732
```

```
               ||||| |||||  || |||||||||| |||||||||| ||||||||||||||||||||
Db         823 GCCCACGTGCTCATGAGCGCCCCAGCGCCGCTTCACTGAGGCTCCATGATGATGTCC 882

Qy         733 CTCACCAAGTTCGCCGACAAGCAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG 792
               |||||||| |||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         883 TTGACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG 942

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAG 852
               ||||| ||||||||||||||||| ||||||||||||||||||||||||| ||| |||||
Db         943 GGCTTCGTGGAGCTCAGCCTGTACGACCAAGTGCGGCTCTTGGAGAGCTGCTGGTTGGAG 1002

Qy         853 GTGTTAATGGCATGGGGCATGATCTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
               ||| | ||| |||||||||||||||||||| || ||||||||||||||||||||||||||
Db        1003 GTGCTCATGGCGTGGCGTAGCATCGACCACCCCGGCAAGCTCATCTTTGCT 1062

Qy         913 CCACATCTTGTTCTGGACAGGGATGAGGCCAAATGCGTACAAGGAATTCTGGAAATCTTT 972
               ||||| || |||||||||||||||| |||||||| || ||||||||||| |||||||||
Db        1063 CCAGACCTGAGTCTGGACAGGCATGAAGCCAAATGCGTTCAAGCAATTCTGAAATCTTT 1122

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||| ||||||||||||| ||||||||||||||||||||||||||||
Db        1123 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATAT 1182

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||| ||||| ||||||||||||||   | |||||  ||
Db        1183 CTCTGTGTCAACGCCATGATCCTGCTCAACTCCAGTATGTACCCTCTCAGCTACAGCATCC 1242

Qy        1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               ||||| ||||||||  |||||||||| ||||| |||||||| |||||||||| || ||| ||
Db        1243 CAGGAGGCTGACAGTGGCCGCAAGCTGACTCACCTGCTGAATGCCGTGACGGACGCTCTG 1302

Qy        1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAAC 1212
               || |||||||||||||||||||||||| |||||||||| || ||||||||||||||||||
Db        1303 GTCTGGGTGATTGCCAAGAGCGGCATGTCCTCCCAGCAACAGTCCATGCGCCTGGCTAAC 1362

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGGAG 1247
               || |||||||||||||| || |||||||||| ||
Db        1363 CTGCTGATGCTCCTGTCTCATGTCAGGCATGCCAG 1397


RESULT 24
AF110402
LOCUS       AF110402               1611 bp    mRNA    linear   MAM 11-APR-1999
DEFINITION  Bos taurus breed Holstein estrogen receptor-beta mRNA, complete
            cds.
ACCESSION   AF110402
VERSION     AF110402.1  GI:4580595
KEYWORDS    .
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1611)
  AUTHORS   Rosenfeld,C.S., Yuan,X., Manikkam,M., Calder,M.D., Garverick,H.A.
            and Lubahn,D.B.
  TITLE     Cloning, sequencing, and localization of bovine estrogen
            receptor-beta within the ovarian follicle
  JOURNAL   Biol. Reprod. 60 (3), 691-697 (1999)
  MEDLINE   99150196
  PUBMED    10026117
REFERENCE   2  (bases 1 to 1611)
  AUTHORS   Yuan,X., Rosenfeld,C.S., Manikkam,M., Calder,M.D., Garverick,H.A.
            and Lubahn,D.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Biochemistry, University of
            Missouri-Columbia, 163A Animal Sciences Research Center, 920 E.
            Campus Drive, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1. .1611
                     /organism="Bos taurus"
                     /strain="breed Holstein"
                     /db_xref="taxon:9913"
```

```
                     /tissue_type "ovarian follicle"
                     /note "RT PCR"
     CDS             28. .1611
                     /function "steroid receptor"
                     /codon_start 1
                     /product "estrogen receptor-beta"
                     /protein_id "AAD24432.1"
                     /db_xref "GI:4580596"
                     /translation "MDVKNSPSSLNSPVSYNCGQSILPEPGPIYLPSSYVESRHEYS
                     AVTFYSPAVMNYSIPNNSEDGPGRQTTSPNVLWPTPGHLSPLAIHCQPSVLYAEPQKS
                     PWRETRSLEHTLPVNRETLKRKASGSSCASPATSPSSKRDAHFCAVCSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKGGSRR
                     ERGGYRIVRRQRNSDEQILHCLSKTKRNGGPMTEVKELLLSALSPEQLVLTLLEAEPPH
                     VLISRPSTPFTFASMMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLILDRDEGKCVEGILEIFDMLLATTSRFRELKLQHR
                     EYLCVKAMILLNSSMYPSATAPQEADSGRKLTHLLNAVTDALVWVIAKSGMSSQQQSM
                     RLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHTLRGNKSLVTGSE
                     RNLVEDSESKEGSQKPQAQ"
BASE COUNT       405 a    428 c    416 g    362 t
ORIGIN

  Query Match           78.9%;  Score 987;  DB 4;  Length 1611;
  Best Local Similarity   87.4%;  Pred. No. 9.7e-214;
  Matches 1080;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy         13 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT 72
              | |  ||| | |  || |||||  |||| ||||||||||||| ||||||||||||||||||
Db        190 AATTACAGTATTCCCAATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGCCCAAAT 249

Qy         73 GTGTTGTGGCCAACACCTGGGCACTTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT 132
              || |||||||||| ||||| ||||| ||||||||||   ||||| |||||   ||| | |
Db        250 GTATTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCAGTT 309

Qy        133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCCCTAGAACACACCTTA 192
              ||||||||  |||||||||||||||||| ||| ||||| | |||||  ||||||||||||||
Db        310 CTGTATGCAGAACCTCAAAAGAGTCCTTGGCGTGAAACCAGATCATTAGAACACACCTTA 369

Qy        193 CCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTT 252
              ||||| |||||  |||||| |||||||||||| ||||||||  | |||| |||||||||| |
Db        370 CCTGTGAACAGGGAGACATTGAAAAGGAAGCCTACTGGGAGCAGTTGTGCCAGCCCTGCT 429

Qy        253 ACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTCGCCTGTCTGCAGCGATTACGCATCG 312
              ||| |||| |||||||||||||||| |||||||| ||||||||||| |||| || |||
Db        430 ACTAGTCCAAGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATCCGTCG 489

Qy        313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 372
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        490 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGTATT 549

Qy        373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGG 432
              |||||||||||||||||||||||||||||||||||||||||||| || |||| |||||
Db        550 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGA 609

Qy        433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGT 492
              || || |||| |||| ||| |||||||||||||| || || || ||||||||||||||||
Db        610 CGTAAAAGCTGTCAGGCCTGTCGACTTCGGAAGTGCTATGAGGTTGGAATGGTGAAGTGT 669

Qy        493 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGACAGAAGTGCCGACGAG 552
              |||| |||||||| || |||||||||||| | |||| || ||||||| | | || |||
Db        670 GGCTCTCGGAGAGAAAGGTGTGGGTACCGCATAGTGCGAAGGCAGAGAAATTCTGATGAG 729

Qy        553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAG 612
              ||||||||||     ||||  ||||||| | || || | |  | ||||||||   |||
Db        730 CAGCTGCACTGCCTGAGCAAAACCAAGAGGAACGGTGGACCCATGACCCGAGTGAAAGAG 789

Qy        613 CTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTCTCACCCTCCTGGAGGCTGAGCCG 672
              |||||||||  |||||||||||| |||||||| ||||| || ||||||||||| ||||||
Db        790 CTGCTGCTGAGCGCCCTGAGCCCAGAGCAGCTGGTCTTACGCTCCTGGAGGCCGAGCCG 849

Qy        673 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 732
              ||||| ||||| || ||||||||||  |||||||||| ||||||||||||||||||||||
Db        850 CCCCACGTGCTCATAAGCCGCCCCAGCACGCCCTTCACTGAGGCCTCCATGATGATGTCC 909

Qy        733 CTGACCAACTTGCCCGACAAGGACTTCGTACACATGATCAGCTGGGCCAAGAAGATCCCG 792
              || |||||| ||||||||||||  |||||||||||||||||||||||||||||||||||
Db        910 CTCACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG 969

Qy        793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAG 852
              ||||| |||||||||||||||| |||||||||||||| |||||||||| ||| |||||
Db        970 GGCTTCGTGGAGCTCAGCCTGTACGACCAAGTGCGGCTTTTGGAGAGCTGCTGGTTGGAG 1029

Qy        853 GTGTTAATGATGGGGCTGATGTGGCGCTCCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
              ||| | ||| |||||||||||||||||||| || ||||||| ||||||||||||||||||
Db       1030 GTGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCTGGCAAGCTCATCTTTGCT 1089

Qy        913 CCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTT 972
              ||||| ||  |||||||||||||||| |||||||| || |||||||||| |||||||||
Db       1090 CCAGACCTCATTCTGGACAGGGATGAAGGAAATGTGTTGAAGGAATTCTAGAAATCTTT 1149

Qy        973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
              |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db       1150 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATAT 1209

Qy       1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
              |||||||||||||||||||||| ||||| |||||||||||||||  | |||||  ||
Db       1210 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTTCAGCTACAGCACCC 1269

Qy       1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
              ||||| |||||||  |||||||||| |||| |||||| || |||| || ||| ||
Db       1270 CAGGAGGCTGACAGTGGCCGGAAGCTGACTCACCTGCTGAATGCTGTGACGGACGCTCTG 1329

Qy       1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAAC 1212
              || ||||||||||||||| ||||| |||||||||||| ||||||||||||||||||||||
Db       1330 GTCTGGGTGATTGCCAAGAGTGGCATGTCCTCCCAGCAGCAGTCCATGCGCCTGGCTAAC 1389

Qy       1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              || |||||||||||| |||||||||| || ||
Db       1390 CTGCTGATGCTCCTGTCTCACGTCAGGCACGCCAG 1424


RESULT 25
BTA18017
LOCUS       BTA18017                1820 bp    mRNA    linear   MAM 09-SEP-1999
DEFINITION  Bos taurus mRNA for estrogen receptor beta.
ACCESSION   Y18017
VERSION     Y18017.1  GI:5823588
KEYWORDS    estrogen receptor beta.
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1820)
  AUTHORS   Walther,N., Lioutas,C., Tillmann,G. and Ivell,R.
  TITLE     Cloning of bovine estrogen receptor beta (ERbeta): expression of
            novel deleted isoforms in reproductive tissues
  JOURNAL   Mol. Cell. Endocrinol. 152 (1-2), 37-45 (1999)
  MEDLINE   99359179
   PUBMED   10432221
REFERENCE   2  (bases 1 to 1820)
  AUTHORS   Walther,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1998) N. Walther, Institute for Hormone and
            Fertility, Research, Grandweg 64, D-22529 Hamburg, FRG
FEATURES             Location/Qualifiers
     source          1. .1820
                     /organism "Bos taurus"
                     /db_xref "taxon:9913"
                     /cell_type "granulosa cells"
     gene            153. .1577
                     /gene "erbeta"
     CDS             153. .1577
                     /gene "erbeta"
                     /codon_start 1
                     /product "estrogen receptor betaa"
                     /protein_id "CAB53861.1"
```

```
                     /db_xref="GI:5823589"
                     /translation="MNYSIPDNSEDGPGRQTTSPNVLWPTPGHLSPLATHCQPSVLYA
                     EPQKSPWRFTRSLEHTLPVNRFTLKRKASGSSCASPATSPSSKRDAHFCAVCSDYASG
                     YHYGVWSCEGCKAFFKRSIQGHNDYTCPATNQCTIDKNRRKSCQACRLRKCYEVGMVK
                     CGSERERCGYRIVPRQRNSDEQTHCLSKTEENQSPMTPVFFLLSAISPQIVTTLLE
                     AEPPHVLISRPSTPFTFASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLE
                     SCWLEVLMVGLMWRSIDHPGKLIFAPDLILDRDEGKCVEGILEIFDMLLATTSRFRFL
                     KLQHKEYLCVKAMILLNSSMYPSATAPQEADSCRKLTHLLNAVTDALVWVIAKSGMSS
                     QQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGNKSI
                     VTGSERNLVEDSESKEGSQKPQAQ"
BASE COUNT      446 a    491 c    478 g    405 t
ORIGIN

  Query Match           78.8%;  Score 985.4;  DB 4;  Length 1820;
  Best Local Similarity   87.4%;  Pred. No. 2.3e-212;
  Matches 1079;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy         13 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGGCCTGGTCGGCAGACCACAAGCCCAAAT 72
              | |  ||| | |  |  ||||| |||| ||||||||||| ||||||||||| ||||||
Db        156 AATTACAGTATTCCCGATAACTCAGAAGATGGGCCTGGTCGACAGACCACAAGTCCAAAT 215

Qy         73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCCCCAGTTATCACAT 132
              || |||||||||| |||| |||| ||||||||||   |||| ||||   |||  |
Db        216 GTATTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCCTCACTT 275

Qy        133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTA 192
              ||||||| |||||||||||||||| ||| |||| | |||| ||||||||||||||
Db        276 CTGTATGCACAACCTCAAAAGAGTCCTTGGCGTGAAACCAGATCATTAGAACACACCTTA 335

Qy        193 CCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTCTT 252
              ||||| |||| |||||||||||||||||||| |||||||| | |||| ||||||||||  |
Db        336 CCTGTGAACAGGGAGACACTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGCT 395

Qy        253 ACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCG 312
              ||| |||| |||||||||||||||| ||||||| |||||||||| |||| || |||
Db        396 ACTAGTCCAAGTTCAAAGAGGCATGCCCACTTCTGTGCTGTCTCCAGTGATTATGCCTGC 455

Qy        313 GGATATCACTATGGAGTCTGGTGCTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 372
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        456 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGTATT 515

Qy        373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGG 432
              |||||||||||||||||||||||||||||||||||||||||| || |||| ||||
Db        516 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACCATAGATAAGAACCGA 575

Qy        433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGCGAATGCTGAAGTCT 492
              || || |||| |||||||| ||||||||||||| || || || ||||||||||||||||
Db        576 CGTAAAAGCTGTCAGGCCTGTCGACTTCGGAAGTGCTATGAGGTTGGAATGCTGAAGTGT 635

Qy        493 GGCTCCCGGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGACAAGTGCCGACGAG 552
              |||||  |||||||| || |||||||||||| | |||| || ||||||| | | || |||
Db        636 GGCTCTCGGAGACAAAGGTGTGGGTACCGCATAGTGCGAAGGCAGAGAAATTCTGATGAG 695

Qy        553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCCCGAGTGCGGGAG 612
              ||||||||||     |||  |||||| |  || || | |  |||||||    |||
Db        696 CAGCTGCACTGCCTGAGCAAAACCAAGAGGAACGGTGGACCCATGACCCGAGTGAAAGAG 755

Qy        613 CTGCTGCTGGAGGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 672
              ||||||||  |||||||||||| |||||||| |||| || ||||||||| ||||||
Db        756 CTGCTGCTGAGCGGCCTGAGCCCAGAGCAGCTGGTGCTTACGCTCCTGGAGGCGGAGCCG 815

Qy        673 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 732
              ||||| |||| || ||||||||||  |||||||||| ||||||||||||||||||||||
Db        816 CCCCACGTGCTCATAAGCCGCCCCAGCACGCCCTTCACTGAGGCCTCCATGATGATGTCC 875

Qy        733 CTGACCAAGTTGGCCGACAAGCAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 792
              || |||||| ||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        876 CTCACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG 935

Qy        793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCCGCCTGTTGGAGAGCTGTTGGATCCTG 852
              ||||| |||||||||||||||| |||||||||||||| |||||||||| ||| |||||
Db        936 GGCTTCGTGGAGCTCAGCCTGTTACGACCAAGTCCGGCTTTTGGAGAGCTGCTGGTTGGAG 995

Qy        853 CTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
              ||| | ||| ||||||||||||||||||||| || |||||||| ||||||||||||||||||
Db        996 CTGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCTGGCAAGCTCATCTTTGCT 1055

Qy        913 CCAGATCTTGTTCTGGACAGGCATGAGGGCAAATGCGTAGAAGGAATTCTGGAAATCTTT 972
              ||||| ||  |||||||||||||||| |||||||| || ||||||||| |||||||||
Db       1056 CCACACCTCATTCTGGACAGGGATGAAGGGAAATGTGTTGAAGGAATTCTAGAAATCTTT 1115

Qy        973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
              |||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||
Db       1116 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAAGAATAT 1175

Qy       1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
              |||||||||||||||||||||| |||| ||||||||||||||   |  ||||  ||
Db       1176 CTCTGTGTCAAGGCCATGATCCTGCTCAACTCCAGTATGTACCCTTCAGCTACAGCACCC 1235

Qy       1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACGGATGCTTTG 1152
              ||||| |||||||| |||||||||| ||||| ||||||| || |||| || ||| ||
Db       1236 CAGGAGGCTGACAGTGGCCGGAAGCTGACTCACCTGCTGAATGCTGTGACGGACGCTCTG 1295

Qy       1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAAC 1212
              || |||||||||||||||| |||| |||||||||||||| ||||||||||||||||||||
Db       1296 GTCTGGGTGATTGCCAAGAGTGGCATGTCCTCCCAGCAGCAGTCCATGCGCCTGGCTAAC 1355

Qy       1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
              || |||||||||||| |||||||||| || ||
Db       1356 CTGCTGATGCTCCTGTCTCACGTCAGGCACGCCAG 1390


RESULT 26
AF164957
LOCUS       AF164957                1581 bp    mRNA    linear   MAM 19-APR-2002
DEFINITION  Sus scrofa estrogen receptor beta mRNA, complete cds.
ACCESSION   AF164957
VERSION     AF164957.2  GI:14717426
KEYWORDS    .
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G., Vale-Cruz,D.S., Choi,I.,
            Katzenellenbogen,B.S., Simmen,F.A. and Simmen,R.C.
  TITLE     Molecular cloning of porcine estrogen receptor beta complementary
            DNAs and developmental expression in periimplantation embryos
  JOURNAL   Biol. Reprod. 66 (3), 760-769 (2002)
  MEDLINE   21858333
   PUBMED   11870084
REFERENCE   2  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G. and Simmen,R.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1999) Animal Science, University of Florida,
            Shealy Drive, Building 459, Gainesville, FL 32611, USA
REFERENCE   3  (bases 1 to 1581)
  AUTHORS   Kowalski,A.A., Graddy,L.G. and Simmen,R.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2001) Animal Science, University of Florida,
            Shealy Drive, Building 459, Gainesville, FL 32611, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 13, 2001 this sequence version replaced gi:5566377.
FEATURES             Location/Qualifiers
     source          1. .1581
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /dev_stage="embryo"
     CDS             1. .1581
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAD45381.2"
                     /db_xref="GI:14717427"
                     /translation="MDIKNSPSNLNSPVSYNCSQSVLPLEPPQPIYLPSSYVESCHEYS
                     AMTFYSPAVVNYSISSNSEVGPGPQATSPNVLWPTPGHLSPLATHCQPSLLYAEPQKS
                     PWCDTRSLEHTLPVNRETLKRKASGSSCASPVTSPSSKRDAHFCAVCSDYASGYHYGV
```

```
                     WSCEGCKAFEKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRR
                     ERCGYRIVRKQRNSEGHLHCLSRAKKNGDHTTRVKELLLSTLSPEQLVLTLLEAEPPH
                     VLVSRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSKFRELKLQHK
                     EYLCVKAMILLNSSMYPSAAAQEAESSRKLTHLLNAVTDALVWVIARSGISSQQQSVR
                     LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHTLRGNKSLVIGSER
                     SRMEESESKEGSQKPQAQ"
BASE COUNT       496 a    424 c    411 g    350 t
ORIGIN

  Query Match             77.0%;  Score 963;  DB 4;  Length 1581;
  Best Local Similarity   86.8%;  Pred. No. 2.6e-207;
  Matches 1072;  Conservative    0;  Mismatches  160;  Indels    3;  Gaps    1;

Qy          14 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT  72
               | |  ||| | |  || ||||  |||| ||||||||||| ||| ||||||||| |||
Db         164 AATTACAGTATTTCCAGTAACTCCGAAGTTGGGCCTGGTCGACAGGCCACAAGCCCCAAT 222

Qy          73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT 132
               ||||||||||||||| ||||| ||||| ||||||||||   |||| |||||   |||| |
Db         223 GTGTTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCCGTCACTT 282

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACACCTTA 192
               ||||||||  |||||||||||||||||| ||||||||  ||||||| ||||||||||||||
Db         283 CTGTATGCAGAACCTCAAAAGAGTCCTTGGTGTGACACAAGATCACTAGAACACACCTTA 342

Qy         193 CCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTGTT 252
               ||||| |||||||||||| |||||||||||| |||||||| | |||| ||||||||||||
Db         343 CCTGTGAACAGAGAGACATTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTGTT 402

Qy         253 ACTGGTCCAGGTTCAAAAGAGGATGCTCACTTCTGCGCCTGTCTGCAGCGATTACGCATCG 312
               ||  ||||| |||||||||||||||| |||||||| |||||||||| ||||| || ||
Db         403 ACCAGTCCAAGTTCAAAAGAGGGATGCCCACTTCTGTGCCTGTCTGCAGTGATTATGCGTCA 462

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 372
               |||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db         463 GGATATCACTATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 522

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACCGG 432
               ||||||||||||||||||||||||||||||||||· |||||||| || ||||| ||||||
Db         523 CAAGGACATAATGATTATATTTGTCCAGCTACAAACCAGTGTACGATAGATAAGAACCGG 582

Qy         433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTGAAGTGT 492
               || || ||||||||||||||||||||||||| ||||| || ||||| ||||||||||||||
Db         583 CGTAAAAGCTGCCAGGCCTGCCGACTTCGAAAGTGCTATGAAGTCGGAATGGTGAAGTGT 642

Qy         493 GGTCCCCGGAGACAGAGATGTGGGTACCGCCTTGTGCCGGAGACAGAGAAGTGCCGACGAG 552
               ||||||||||||  | |||||||||||| | ||||| |  ||||||| | | || | |
Db         643 GGCTCCCCGGAGAGAACGGTGTGGGTACCGCCATAGTGCGAAAGCAGAGAAATTCTGAGGGG 702

Qy         553 CAGCTGCACTGTGCCGTAAGGCCAAGAGAAGTGGCCGCCACGCGCCCCGAGTGCGGGAG 612
               || ||||||||     ||| |||||||  | ||| |  ||| |  |||||||  ||||
Db         703 CATCTGCACTGCCTGAGCAGAGCCAAGAAGAATGGTGATCACACAACCGGAGTGAAGGAG 762

Qy         613 CTGCTGCTGGAGGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTGGAGGCTGAGCCG 672
               ||||| ||| | ||||||||||| |||||||| |||||||| ||||||||||| |||||
Db         763 CTGCTCCTGAGCACCCTGAGCCCAGAGCAGCTGGTGCTCACGCTCCTGGAGGCCGAGCCA 822

Qy         673 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATGATGTCC 732
               ||||| |||||| | ||||||||||  ||||||||||||||||||||||||||||||||
Db         823 CCCCACGTGCTGGTGAGCCGCCCCAGCACGCCCTTCACCGAGGCCTCCATGATGATGTCC 882

Qy         733 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCC 792
               |||||||| ||||||||||||  ||||||||||||||||||||||||||| |||||
Db         883 CTGACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAAATTCCC 942

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCCTGGAGAGCTGCTGGTTGCAG 852
               |||||  |||||||||||||||| ||||||||||||||| |||||||||| ||| |||||
Db         943 GGCTTCATGGAGCTCAGCCTGTACGACCAAGTGCGGCTCCTGGAGAGCTGCTGGTTGCAG 1002

Qy         853 GTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
               ||| | ||| ||||||||||||||||| || ||||||||||||||||||||||||||||||
Db        1003 GTGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCT 1062

Qy         913 CCAGATCTTGTTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATTCTGGAAATCTTT 972
               ||||| || |||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db        1063 CCAGACCTGTTTCTGGACAGGGATGAAGGCAAATGCGTAGAAGGAATTCTGGAAATCTTT 1122

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||||| |||||||||||||| |||||||||||||||||||| || |||
Db        1123 GACATGCTCCTGGCAACGACTTCAAGGTTTCGTGAGTTAAAACTCCAACACAAGGAGTAT 1182

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||||||| ||||| |||||||||||| |||   |  | ||  ||
Db        1183 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTATCCTTCAG-  CTGCAGCC 1239

Qy        1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               || || ||||| |||||||||||||||| ||||  ||||||| |||||||| ||||| |||
Db        1240 CAAGAGGCTGAGAGCAGCCGGAAGCTGACTCATCTGCTGAATGCCGTGACTGATGCCTTG 1299

Qy        1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCGCCTGGCTAAC 1212
               || |||||||| |||| ||| || ||||||||||||||||||| ||  | ||||||| |||
Db        1300 GTCTGGGTGATCGCCAGGAGTGGTATCTCCTCCCAGCAGCAGTCTGTCCGCCTGGCCAAC 1359

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               ||||| ||||||| || || ||||| ||||| ||
Db        1360 CTCCTCATGCTCCTTTCTCATGTCAGACATGCCAG 1394


RESULT 27
AF267736
LOCUS       AF267736                1581 bp    mRNA    linear   MAM 15-MAR 2002
DEFINITION  Sus scrofa estrogen receptor beta (ER beta) mRNA, complete cds.
ACCESSION   AF267736
VERSION     AF267736.1  GI:13183320
KEYWORDS    .
SOURCE      Sus scrofa.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1581)
  AUTHORS   LaVoie,H.A., DeSimone,D.C., Gillio-Meina,C. and Hui,Y.Y.
  TITLE     Cloning and characterization of porcine ovarian estrogen receptor
            beta isoforms
  JOURNAL   Biol. Reprod. 66 (3), 616-623 (2002)
  MEDLINE   21858315
   PUBMED   11870066
REFERENCE   2  (bases 1 to 1581)
  AUTHORS   LaVoie,H.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY 2000) Cell Biology and Neuroscience, University
            of South Carolina, School of Medicine, Bldg 1, Columbia, SC 29208,
            USA
FEATURES             Location/Qualifiers
     source          1..1581
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /tissue_type="ovary"
     gene            1..1581
                     /gene="ER beta"
     CDS             1..1581
                     /gene="ER beta"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAK15151.1"
                     /db_xref="GI:13183321"
                     /translation="MDIKNSPSNLNSPVSYNCSQSVLPLEPGPIYIPSSYVESCHEYS
                     AMTFYSPAVVNYSISSNSEVGPGRQATSPNVLWPTPGHLSPLATHCQPSLLYAEPQKS
                     PWCDIRSLEHLPVNKELLKKKASGSSCASPVTSPSSKRDAHFCAVGSDYASGYHYGV
                     WSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCGSRR
                     ERCGYRIVRKQRNSEGHLHCLSRAKKNGDHTTRVKELLLSTLSPEQLVLTLLEAEPPH
                     VLVSRPSTPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLYDQVRLLESCWLE
                     VLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSKFRELKLQHK
                     EYLCVKAMILLNSSMYPSAAAQEAESSRKLTHLLNAVTDALVWVIARSGISSQQQSVR
                     LANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHTLRGNKSLVIGSER
                     SRMEESESKEGSQKPQAQ"
```

```
BASE COUNT      394 a    425 c    412 g    350 t
ORIGIN

  Query Match             77.0%;  Score 963;  DB 4;  Length 1581;
  Best Local Similarity   86.9%;  Pred. No. 2.6e-207;
  Matches 1072;  Conservative    0;  Mismatches  160;  Indels    3;  Gaps    1;

Qy          13 ATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACCACAAGCCCAAAT 72
               | |  ||| | |  || ||||  |||| ||||||||||| ||| |||||||||| |||
Db         163 AATTACAGTATTTCCAGTAACTCCGAAGTTGGGCCTGGTCGACAGGCCACAAGCCCCAAT 222

Qy          73 GTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGCCAGTTATCACAT 132
               |||||||||||||| ||||| ||||| |||||||||   ||||| |||||   |||| |
Db         223 GTGTTGTGGCCAACTCCTGGACACCTCTCTCCTTTAGCCATCCATTGCCAGCGGTCACTT 282

Qy         133 CTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTAGAACACAGCTTA 192
               ||||||| |||||||||||||||| |||||||  ||||||| ||||||||||||||||
Db         283 CTGTATGCAGAACCTCAAAAGAGTCCTTGGTGTGACACAAGATCACTAGAACACAGCTTA 342

Qy         193 CCTGTAAACAGAGAGACACTGAAAACGAAGGTTAGTGGGAACCGTTGCGCCAGCCCTCTT 252
               ||||| |||||||||||| |||||||||| |||||||| | |||| ||||||||||||
Db         343 CCTGTGAACAGAGAGACATTGAAAAGGAAGGCTAGTGGGAGCAGTTGTGCCAGCCCTCTT 402

Qy         253 ACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGCGATTACGCATCG 312
               ||  ||||| |||||||||||||||| |||||||| |||||||||| ||||| || ||
Db         403 ACCAGTCCAAGTTCAAAGAGGGATGCCCACTTCTGTGCTGTCTGCAGTGATTATGCGTCA 462

Qy         313 GGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 372
               ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         463 GGATATCACTATGGAGTCTGGTCATGTGAAGGATGTAAGGCCTTTTTTAAAAGAAGCATT 522

Qy         373 CAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATCGATAAAAACGGG 432
               |||||||||||||||||||||||||||||||||||| |||||||| || ||||| |||||
Db         523 CAAGGACATAATGATTATATTTGTCCAGCTACAAACCAGTGTACGATAGATAAGAACGGG 582

Qy         433 CGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGAATGGTCAAGTGT 492
               || || ||||||||||||||||||||||| ||||| || ||||| ||||||||||||||
Db         583 CGTAAAAGCTGCCAGGCCTGCCGACTTCGAAAGTGCTATGAAGTCGGAATGGTCAAGTGT 642

Qy         493 GGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGACAAGTGCCGACGAG 552
               |||||||||||||  | |||||||||||| | ||||| |  ||||||| | || || | |
Db         643 GGCTCCCGGAGAGAACGGTGTGGGTACCGCATAGTGCGAAAGCAGACAAATTCTGACGGG 702

Qy         553 CAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGCCCACGCGCCCCGAGTGCCGGAG 612
               || ||||||||     |||  |||||||  | ||| |  ||| |  ||||||||  ||||
Db         703 CATCTGCACTGCCTGAGCAGAGCCAAGAAGAATGGTGATCACACAACCCGAGTGAAGGAG 762

Qy         613 CTGCTGCTGGACGCCCTGAGCCCGCAGCAGCTACTGCTGACCCTCCTGGAGGTTGAGGGG 672
               ||||| |||  | ||||||||||  |||||||| |||||||| |||||||||| |||||
Db         763 CTGCTCCTGAGCACCCTGAGCCCAGCAGCAGCTGGTGCTCACGCTCCTGCAGGCCGAGCCA 822

Qy         673 CCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCCATGATCATGTCC 732
               ||||| |||||| | ||||||||||  ||||||||||||||||||||||||| |||||
Db         823 CCCCACGTGCTGGTGAGCCGCCCCAGCACGCCCTTCACCGAGGCCTCCATGATCATGTCC 882

Qy         733 CTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCCAAGAAGATTCCG 792
               ||||||||| |||||||||||||  |||||||||||||||||||||||| |||||
Db         883 CTGACCAAGCTGGCCGACAAGGAACTGGTACACATGATCAGCTGGGCCAAGAAAATTCCG 942

Qy         793 GGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGCTGTTGGATGGAG 852
               ||||| |||||||||||||| |||||||||||||| |||||||||| ||| |||||
Db         943 GGCTTCGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCCTGGAGAGCTGTTGCTTCCAC 1002

Qy         853 GTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAGCTCATCTTTGCT 912
               ||| | ||| |||||||||||||||||||| || ||||||||||||||||||||||||
Db        1003 GTGCTCATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAGCTCATCTTTGCT 1062

Qy         913 CCAGATCTTGTTCTGGACAGGGATGAGGGAAATGCGTAGAAGGAATTCTGGAAATCTTT 972
               ||||| || |||||||||||||||| |||||||| ||||||||||||||||||||||||
Db        1063 CCAGACCTCGTTCTGGACAGGGATGAAGGGAAATGTGTAGAAGGAATTCTGGAAATCTTT 1122

Qy         973 GACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAACACAAAGAATAT 1032
               |||||||||||||||| |||||||||||| ||||||||||||||||||||| || |||
Db        1123 GACATGCTCCTGCCAACGACTTCAAGGTTTCCTCAGTTAAAACTCCAACACAAGGAGTAT 1182

Qy        1033 CTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTGGTCACAGCGACC 1092
               |||||||||||||||||||||||| ||||| |||||||||| |||   |  | ||  ||
Db        1183 CTCTGTGTCAAGGCCATGATCCTCCTCAACTCCAGTATGTATCCTTGACA---CTGCAGCC 1239

Qy        1093 CAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTGACCGATGCTTTG 1152
               || || |||| |||||||||||||| ||||  ||||||| ||||||| |||||  ||
Db        1240 CAAGAGGCTGAGAGCAGCCGGAAGCTGACTCATCTGCTGAATGCCGTGACTGATGCCCTG 1299

Qy        1153 GTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATGCCCCTGCCTAAC 1212
               || |||||||| |||| ||| || ||||||||||||||||| ||  | |||||||| |||
Db        1300 GTCTGGGTGATGGCCAGGAGTGGTATCTCCTCCCAGCAGCAGTCTGTCCGCCTGGCCAAC 1359

Qy        1213 CTCCTGATGCTCCTGTCCCACGTCAGGCATGCGAG 1247
               ||||| |||||||| || || ||||| |||| ||
Db        1360 CTCCTCATGCTCCTTTCTCATGTCAGACATGCCAG 1394


RESULT 28
AF124790
LOCUS       AF124790               1427 bp    mRNA    linear   PRI 27-MAY-1999
DEFINITION  Homo sapiens estrogen receptor beta2 splice variant mRNA, complete
            cds.
ACCESSION   AF124790
VERSION     AF124790.1  GI:4894663
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1427)
  AUTHORS   Lu,B., Leygue,E., Dotzlaw,H., Watson,P.H. and Murphy,L.C.
  TITLE     Cloning and functional characteristics of human estrogen receptor
            beta isoforms
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1427)
  AUTHORS   Lu,B., Leygue,E., Dotzlaw,H., Watson,P.H. and Murphy,L.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Biochemistry, University of Manitoba, 770
            Bannatyne Avenue, Winnipeg, MB R3E 0W3, Canada
FEATURES             Location/Qualifiers
     source          1. .1427
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="ovary"
     CDS             52. .1023
                     /note "missing exon 5"
                     /codon_start 1
                     /product "estrogen receptor beta2 splice variant"
                     /protein_id="AAD32580.1"
                     /db_xref="GI:4894663"
                     /translation="MDIKNSPSSILNSPSSYNCSQSILPLEHGSIYIPSSYVDSHHEYP
                     AMTFYSPAVMNYSIPSNVTNLEGGPGRQTTSPNVLWPTPGHLSPLVVHRQLSHLYAEP
                     QKSPWCEARSLEHTLPVNRETLKRKVSGNRCASPVTGPGSKRDAHFCAVCSDYASGYH
                     YGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLRKCYEVGMVKCG
                     SPRERCGYRLVRRQRSADFQLHCAGKAKRSGCHAPPVRELLIDAISPFQLVLILLEAE
                     PPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGMPGNA"
BASE COUNT      357 a    384 c    464 g    322 t
ORIGIN

  Query Match             76.7%;  Score 960;  DB 9;  Length 1427;
  Best Local Similarity   88.9%;  Pred. No. 1.2e-206;
  Matches 1109;  Conservative    0;  Mismatches    0;  Indels  139;  Gaps    1;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 ATGAATTACAGCATTCCCAGCAATCTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 270

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 330
```

```
Qy       121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       331 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 390

Qy       181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       391 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 450

Qy       241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       451 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 510

Qy       301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       511 GATTACGCATCGGGATATCACTATGCAGTCTGGTCGTGTCAAGCATGTAAGGCCTTTTTT 570

Qy       361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       571 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 630

Qy       421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       631 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 690

Qy       481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       691 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 750

Qy       541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGCGCC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       751 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGCGCC 810

Qy       601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       811 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 870

Qy       661 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       871 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 930

Qy       721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       931 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 990

Qy       781 AAGAAGATTCCCCGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
            ||||||||||||
Db       991 AAGAAGATTCCC- - - -
```

```
Db        510 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 569

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 629

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGCAAGGTTAGTGGGAACCGTTGC 689

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 749

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGCATGTAAGGCCTTTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGCATGTAAGGCCTTTTTT 809

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 869

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 929

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 ATGGTGAAGTGTGGCTCCCGGAGACAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 989

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACCGCCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGCCGGCCACCGCCGC 1049

Qy        601 CGAGTGCGGGCAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 CGAGTGCGGGCAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 1109

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 1169

Qy        721 ATGATGATGTCGCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 ATGATGATGTCGCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 1229

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||||||||||
Db       1230 AAGAAGATTCCC---------------------------------------------------- 1241

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
                                                                          |||
Db       1242 ------------------------------------------------------------ 1241

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
                                                 ||||||||||||||||||||||||||||||
Db       1242 ---------------------------------GGGATGAGGGGAAATGCGTAGAAGGAATT 1270

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1271 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1330

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCGTCTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1331 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCGTCTG 1390

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1391 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1450

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1451 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATC 1510

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATCGGAGG 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1511 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCATCGGAGG 1558


RESULT 30
AF042058
LOCUS       AF042058                1621 bp    mRNA    linear   ROD 22-JAN-1998
DEFINITION  Rattus norvegicus estrogen receptor beta1 (ERbeta) mRNA, complete
            cds.
ACCESSION   AF042058
VERSION     AF042058.1  GI:2801690
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1621)
  AUTHORS   Petersen,D.N., Tkalcevic,G.T., Koza-Taylor,P.H., Turi,T.G. and
            Brown,T.A.
  TITLE     Identification of estrogen receptor beta2, a functional variant of
            estrogen receptor beta expressed in normal rat tissues
  JOURNAL   Endocrinology (1998) In press
REFERENCE   2  (bases 1 to 1621)
  AUTHORS   Brown,T.A. and Petersen,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1998) Molecular Sciences, Pfizer Central
            Research, Eastern Point Road, Groton, CT 06340, USA
FEATURES             Location/Qualifiers
     source          1. .1621
                     /organism "Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .1621
                     /gene="ERbeta"
     CDS             1. .1458
                     /gene="ERbeta"
                     /note="steroid hormone receptor"
                     /codon_start=1
                     /product="estrogen receptor beta1"
                     /protein_id="AAB97424.1"
                     /db_xref="GI:2801691"
                     /translation="MTFYSPAVMNYSVPDSTSNLDGDVRQSTSPNVIWPTSGHLSPL
                     ATHCQSSLLYAEPQKSPWCEARSLEHTLPVNRETLKRKLSGSSCASPVTSPNAKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LRKCYEVGMVKGGSRRERPCGYRIVPRQRSSSFQVHCISKAKRNGCHAPRVKFLLISIL
                     SPEQLVLILEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLTHLLNAVTDAL
                     VWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHTLRGYKSSISGSECSSTEDSKNKESSQNLQSQ"
     variation       80
                     /gene="ERbeta"
                     /note="encodes leucine residue in GenBank Accession Number
                     U57439"
                     /replace="t"
     variation       813
                     /gene="ERbeta"
                     /note="encodes proline residue in GenBank Accession Number
                     U57439"
                     /replace="c"
BASE COUNT      404 a    417 c    449 g    351 t
ORIGIN

  Query Match           75.2%;  Score 940.2;  DB 10;  Length 1621;
  Best Local Similarity   84.9%;  Pred. No. 3.8e-202;
  Matches 1054;  Conservative    0;  Mismatches  188;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  ||||| |||    || |||| |||| ||||||||||  || |||| |
Db         25 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGAGGGTGGGCCTGTCCGACAGAGC 84
```



```
Qy        61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
             ||||||||||||||| | ||||||||  ||||||||| ||||||||| |  |||| ||
Db        85 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 144

Qy       121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
             || | ||| |  || ||||| ||||||||||||||||| ||||||||||||||||| |||
Db       145 CAATCATGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy       181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
             || ||||||||||||||||||||||||||||||||| |||||| |||||||| | ||||
Db       205 GAGCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCAGTTGT 264

Qy       241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
             |||||||||||||||| |||||    |||||||||||||||||||||||| |||||||||
Db       265 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCGCCGTCTGCAGC 324

Qy       301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
             ||||| |||| || |||| || || || |||| ||||||||||||||||||||||||||
Db       325 GATTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTT 384

Qy       361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
             |||||||||||||||||||||||||||||||| ||||||| || |||||||||| ||
Db       385 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 444

Qy       421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
             || || |||||||| || ||||||||||||||||||||||| |||||||| |||| |||
Db       445 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 504

Qy       481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
             ||||| ||||||| ||| ||||||| | |||||||||| | |||||||| |||||
Db       505 ATGGTCAAGTGTGGATCCAGGAGAGAAGGGTGTGGGTACCGTATAGTGCGGAGGCAGAGA 564

Qy       541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGGCCC 600
             ||| |  |||||| | ||||  |  |||| |||||||||  || || || || |||
Db       565 AGTTCTAGCGAGCAGCTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCC 624

Qy       601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTGCTG 660
             || |||  |||||| ||||||  | || |||| || |||||  || ||||||||||||||
Db       625 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTGCTG 684

Qy       661 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
             || ||||| || ||| |||||||| | ||||| ||||   |||||||||||||||||||
Db       685 GAAGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy       721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
             |||||||||||||| || ||| |||| |||||||  |||| ||||||||  |||||||
Db       745 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy       781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
             ||||| || || |||||||||||||||||| |||||||| |||||||| || |||
Db       805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGC 864

Qy       841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
             || |||||||||| ||||| |||| |||||||||||||| || ||||||||||||||||
Db       865 TGCTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy       901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
             ||||| || |||| || || ||||||||||||||||||||||| |||||||||| |||
Db       925 CTCATTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGGATT 984

Qy       961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
             |||||||||||||||| |||||||| || || |||||||| || ||||||||||||||
Db       985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1044

Qy      1021 CACAAAGAATATCTCTTGTGTCAAGGCCATGATCCTCCTCAATTCCAGTATGTACCCTTG 1080
             ||||| || |||||||||||| ||||||||||||| ||||| ||||||||||||||  ||
Db      1045 CACAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCCTTG 1104

Qy      1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCGTG 1140
             |   | || | |||||| || || || |||||||||||| | ||| | |||||||| |||
Db      1105 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1164

Qy      1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
             || ||||  |||| |||||||||| |||| || ||||||||||||||||| ||  |
Db      1165 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1224

Qy      1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCA 1241
             || |||| |||||||||||||| || || ||||||||||
Db      1225 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCA 1265


RESULT 31
RNAJ2602
LOCUS       RNAJ2602                1650 bp    DNA     linear   ROD 18-APR-1998
DEFINITION  Rattus norvegicus mRNA for estrogen receptor beta (clone
            pcDNA.ERbeta).
ACCESSION   AJ002602
VERSION     AJ002602.1  GI:3077648
KEYWORDS    estrogen receptor beta.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV 1997) Aldridge T.C., Molecular Endocrinology
            Group, ZENECA Central Toxicology Laboratory, Alderley Park,
            Macclesfield, Cheshire SK10 4TJ, ENGLAND
  REMARK    revised by [3]
REFERENCE   2  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Tissue specific responses to estrogen: an explanation based on
            differential activation of multiple estrogen receptors with
            different estrogen response elements
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1650)
  AUTHORS   Aldridge,T.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1997) Aldridge T.C., Molecular Endocrinology
            Group, ZENECA Central Toxicology Laboratory, Alderley Park,
            Macclesfield, Cheshire SK10 4TJ, ENGLAND
COMMENT     On Apr 23, 1998 this sequence version replaced gi:2642166.
FEATURES             Location/Qualifiers
     source          1..1650
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone="pcDNA.ERbeta"
                     /tissue_type="Prostate"
     gene            1..1650
                     /gene="ERbeta"
     CDS             1..1650
                     /gene="ERbeta"
                     /note="The clone was isolated by RT-PCR from total RNA"
                     /codon_start=1
                     /evidence=experimental
                     /product="estrogen receptor beta"
                     /protein_id="CAA05641.1"
                     /db_xref="GI:3077649"
                     /db_xref="SPTREMBL:O35784"
                     /translation="MSICTSSHKEESQLRQSSDMETKNSPSSLSSPASYNCSQSTLPL
                     EHGPIYIPSSYVDNRHEYSAMTFYSPAVMNYSVPGSTSNLDGGPVRQSTSPNVLWPTS
                     GHLSPLATHCQSSLIYAEPQKSPWCEAKSIFHTTPVNREFLTKPKTSGSSCASPVTSPN
                     AKRDAHFCAVCSDYASGYHYGVWPCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRK
                     SCQACRLRKCYEVGMVKCGSRRERCGYRIVRRQRSSSEQVHCLSKAKRNGGHAPRVKE
                     LITSIISPEQIVLTLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKK
                     IPGFVELSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGI
                     LEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQEAESSRKLTHLLN
                     AVTDALVWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYD
                     LLLEMLNAHVLRGYKPSISSGSEGSSIEDSKNKESSQNLQSQ"
BASE COUNT      415 a     443 c     432 g     360 t
ORIGIN
```

```
  Query Match             75.2%;  Score 940.2;  DB 10;  Length 1650;
  Best Local Similarity   84.9%;  Pred. No. 3.8e-202;
  Matches 1053;  Conservative    0;  Mismatches  188;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| ||||| ||||| ||| || ||| |||| |||||||||| || |||| |
Db        217 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACAGAGC 276

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | |||||||| |||||||| ||||||||| | |||| ||
Db        277 ACAAGCCCAAATGTGCTATGCCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 336

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAACAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| | || |||||||||||||||| |||||||||||||| ||| |
Db        337 CAATCATCGCTCCTCTATGCAGAACCTCAAAACAGTCCTTGGTGTGAAGCAAGATCACTA 396

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||||| |||||| |||||||| | ||||
Db        397 GAGCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCCAGTGT 456

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||| |||||    |||||||||||||||||||||||| ||||||||
Db        457 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCGCCGTCTGCCAGC 516

Qy        301 GATTACGCATCGGGATATCACTATGCAGTCTGGTCGTGTGAAGGATCTAAGGCCTTTTTT 360
              ||||| |||| || |||| || || || || | ||||||||||||||||||||||||||
Db        517 GATTATGCAICTGGGTATCATTACGGCGTTTGGCCATGTGAAGGATCTAAGGCCTTTTTT 576

Qy        361 AAAACAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||||||| |||||||| || |||||||||| ||
Db        577 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 636

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| || ||||||||||||||||||||||| |||||||| ||||| |||
Db        637 GACAAGAACCGGCGTAAAAGCTGCCAGGCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 696

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGCAGACAGAGA 540
              ||||| ||||||| ||| |||||||  | ||||||||||  | |||||||| ||||||
Db        697 ATGGTCAAGTGTGGATCCAGGAGAGAACGGTGTGGGTACCGTATAGTGCGGAGCCACAGA 756

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGCCAAGGCCAAGAGAAGTGCGCGCCACGCGCCC 600
              ||| |   ||||||| | |||||     |||| ||||||||||  || || ||||| |||
Db        757 AGTTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCACGCACCC 816

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTACTGCTCACCCTCCTG 660
              || |||  |||||| ||||||  | || |||| || |||| || ||||||||||||||||
Db        817 CGGGTGAAGCAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 876

Qy        661 GAGGCTGAGCCCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGCCCTCC 720
              || ||||| || ||| |||||||| | |||| ||||   ||||||||||||||||||||
Db        877 GAAGCTGAACCCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGCCCTCC 936

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| ||||||||  |||| ||||||||  ||||||||
Db        937 ATGATGATGTCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATCATTGGCTGGGCC 996

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCCTGGAGAGC 840
              ||||| || |||||||||||||||||||||||| ||||||| ||||||| || |||
Db        997 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTCCGGCTCCTAGAAAGC 1056

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGTTGATGTGGGGCTCAATTGACCACCCCCCTAAC 900
              || |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||
Db       1057 TGCTGGATGGAGGTGCTAATGGTGGGACCGATGTGGCGCTCCATCGACCACCCCGGCAAG 1116

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGCATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||| || ||||| || || |||||||||||||||||||||| |||||||||| |||
Db       1117 CTCATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGGATT 1176

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGACACTTAAAACTCCAA 1020
              |||||||||||||||||||||||||||| || || |||||||| || ||||||||||||
Db       1177 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1236

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||| || |||||||||| |||||||||||| ||||||||||| |||| |||||||||||  ||
Db       1237 CACAAGGACTATCTCTCTGTCAAGCCCATGATCCTCCTCAACTCCAGTATGTACCCTCTG 1296

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |   | || | ||||| || || || |||||||||||| | ||| | ||||||| |||
Db       1297 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1356

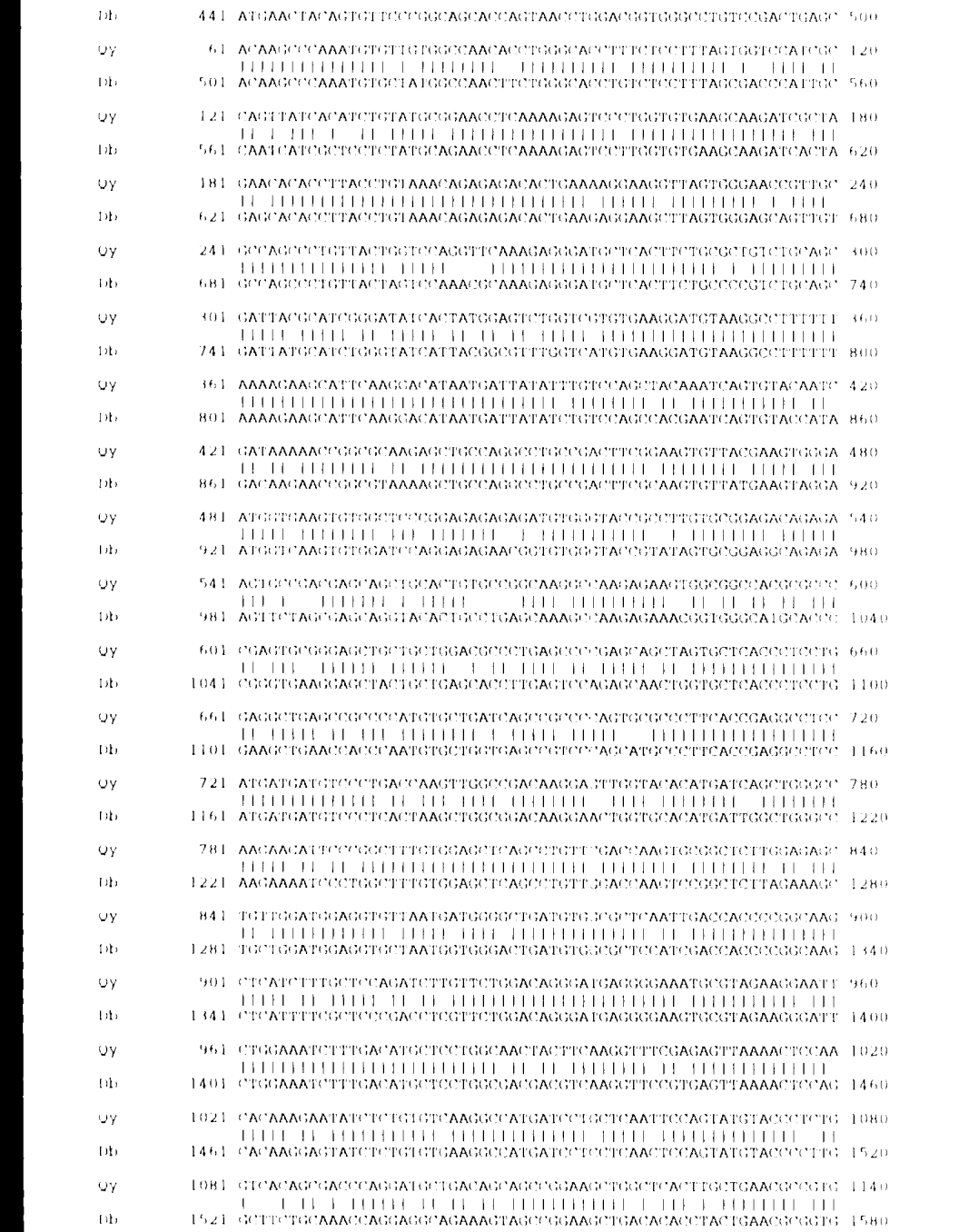
Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||| |||| |||||||||| |||| || |||||||||||||||| ||  |
Db       1357 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1416

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCA 1241
              || |||| |||||||||||||| || || ||||||||||
Db       1417 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCA 1457


RESULT 32
RNU57439
LOCUS       RNU57439                2555 bp    mRNA    linear   ROD 12-JUN-1996
DEFINITION  Rattus norvegicus estrogen receptor beta mRNA, complete cds.
ACCESSION   U57439
VERSION     U57439.1  GI:1373280
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 8 to 2555)
  AUTHORS   Kuiper,G.G., Enmark,E., Pelto-Huikko,M., Nilsson,S. and
            Gustafsson,J.A.
  TITLE     Cloning of a novel receptor expressed in rat prostate and ovary
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (12), 5925-5930 (1996)
  MEDLINE   96234066
   PUBMED   8650195
REFERENCE   2  (bases 1 to 2555)
  AUTHORS   Enmark,E.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-1996) Eva L. Enmark, Center for Biotechnology,
            NOVUM, Karolinska, Institute, NOVUM, Huddinge, S-141 57, Sweden
FEATURES             Location/Qualifiers
     source          1. .2555
                     /organism "Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /sex="male"
                     /tissue_type "prostate"
     CDS             417. .1874
                     /note="nuclear receptor superfamily member"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id "AAC52602.1"
                     /db_xref "GI:1373281"
                     /translation="MTFYSPAVMNYSVPGSESNIDGGPVRLSTSPNVIWPTSGHLSPL
                     ATHCQSSLLYAEPQKSPWCEAKSLEHTLPVNPETLEPKLSGSSCASPVTSPNAKRDAH
                     FCPVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LRKCYEVGMVKGGSPRERCGYRHVPQPSSSPQVHCLSKAKPNGGHAPRVKELLLSTL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLASANQPAESSRKLIHLLNAVTDAL
                     VWVIAESCISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHTLRGYKSSISGSECSSTEDSKNKESSQNLQSQ"
     misc_feature    726. .923
                     /note="encodes DNA binding domain"
BASE COUNT      630 a    652 c    660 g    613 t
ORIGIN

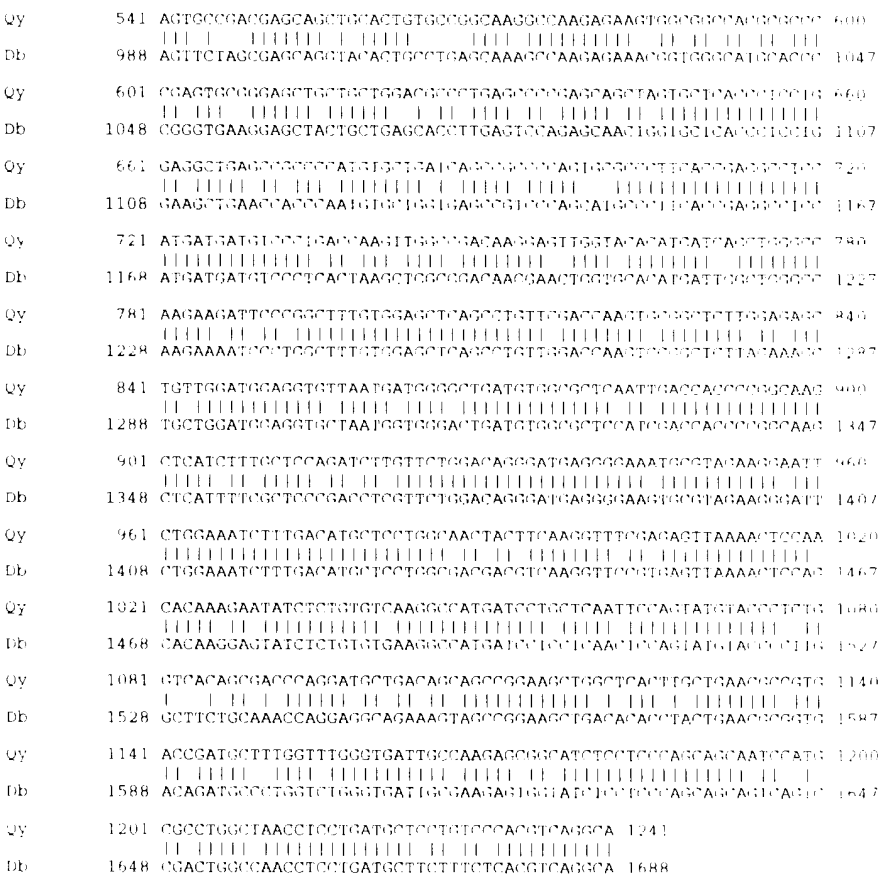
  Query Match             74.9%;  Score 937;  DB 10;  Length 2555;
  Best Local Similarity   84.7%;  Pred. No. 2e-201;
  Matches 1051;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| ||||| ||||| ||| || ||| |||| |||||||||| || | || |
```



RESULT 33
A61583
LOCUS A61583 2568 bp DNA linear PAT 09-MAR-1998
DEFINITION Sequence 1 from Patent WO9709348.
ACCESSION A61583
VERSION A61583.1 GI:3715916
KEYWORDS .
SOURCE Rattus rattus.
ORGANISM Rattus rattus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2568)
AUTHORS Kuiper,G.G., Enmark,E. and Gustafsson,J.
TITLE ORPHAN RECEPTOR
JOURNAL Patent: WO 9709348-A 1 13-MAR-1997;
KAROBIO AB (SE)
FEATURES Location/Qualifiers
source 1..2568
/organism "Rattus rattus"
/db_xref "taxon:10117"
BASE COUNT 634 a 654 c 663 g 617 t
ORIGIN

Query Match 74.9%; Score 937; DB 6; Length 2568;
Best Local Similarity 84.7%; Pred. No. 2e-201;
Matches 1051; Conservative 0; Mismatches 190; Indels 0; Gaps 0;



```
RESULT 34
AR075920
LOCUS       AR075920               2568 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 1 from patent US 5958710.
ACCESSION   AR075920
VERSION     AR075920.1  GI:10002666
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2568)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J.-A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710-A 1 28-SEP-1999.
FEATURES             Location/Qualifiers
     source          1..2568
                     /organism "unknown"
BASE COUNT      634 a    654 c    663 g    617 t
ORIGIN

  Query Match             74.9%;  Score 937;  DB 6;  Length 2568;
  Best Local Similarity   84.7%;  Pred. No. 2e-201;
  Matches 1051;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;
```



```
Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |   | || | |||||| || || || ||||||||||||| | ||| | ||||||| |||
Db       1528 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1587

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||||| ||||| || ||||||||||||||||||| ||  |
Db       1588 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1647

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCA 1241
              || ||||| |||||||||||||| || || ||||||||||
Db       1648 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCA 1688


RESULT 45
AX022455
LOCUS       AX022455               2568 bp    DNA     linear   PAT 20-SEP 2000
DEFINITION  Sequence 1 from Patent EP0935000.
ACCESSION   AX022455
VERSION     AX022455.1  GI:10046068
KEYWORDS    .
SOURCE      black rat.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae;
            Rattus.
FEATURES             Location/Qualifiers
     source          1..2568
                     /organism="Rattus rattus"
                     /db_xref="taxon:10117"
BASE COUNT      644 a    654 c    663 q    617 t
ORIGIN

  Query Match             74.9%;  Score 937;  DB 6;  Length 2568;
  Best Local Similarity   84.7%;  Pred. No. 2e-201;
  Matches 1051;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAAGTTCGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| ||||| |||||| ||    || |||| |||| ||||||||||  || | || |
Db        448 ATGAACTACAGTGTTCCCGGCAGCACCAGTAACCTGGACGGTGGGCCTGTCCGACTGAGC 507

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | |||||||  |||||||||| |||||||| |  |||| ||
Db        508 ACAAGCCCAAATGTGCTATGGCCAACTTCTGGGCACCTGTCTCCTTTAGCGACCCATTGC 567

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| |  || ||||| |||||||||||||||| |||||||||||||||| |||
Db        568 CAATCATGGCTCCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 627

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              || |||||||||||||||||||||||||||||| |||||| |||||||| | ||||
Db        628 GAGCACACCTTACCTGTAAACAGAGAGACACTGAAGAGGAAGCTTAGTGGGAGCCACTTCT 687

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||| |||||    |||||||||||||||||||||| | |||||||||
Db        688 GCCAGCCCTGTTACTAGTCCAAACGCAAAGAGGGATGCTCACTTCTGCCCCGTCTGCAGC 747

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || |||| || || || |||| ||||||||||||||||||||||||||
Db        748 GATTATGCATCTGGGTATCATTACGGCGTTTGGTCATGTGAAGGATGTAAGGCCTTTTTT 807

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              ||:|||||||||||| |||||||||||||| |||||||| || |||||||||| ||
Db        808 AAAAGAAGCATTCAAGGACATAATGATTATATCTGTCCAGCCACGAATCAGTGTACCATA 867

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| || |||||||||||||||||||||||| |||||||| |||| |||
Db        868 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTATGAAGTAGGA 927

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGACACACAGA 540
              ||||| |||||||| ||| ||||||| | |||||||||||  | |||||||| |||||
Db        928 ATGGTCAAGTGTGGATCCAGGAGAGAACGCTGTGGGTACCGTATAGTGCGGAGGCAGAGA 987

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
```

```
              ||| |  |||||||| | |||||     |||| |||||||||| || || || || |||
Db        988 AGTTCTAGCGAGCAGGTACACTGCCTGAGCAAAGCCAAGAGAAACGGTGGGCATGCACCC 1047

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  ||||| ||||||  | || |||| || |||| || ||||||||||||||||
Db       1048 CGGGTGAAGGAGCTACTGCTGAGCACCTTGAGTCCAGAGCAACTGGTGCTCACCCTCCTG 1107

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || ||||| || ||| |||||||| | |||| |||||   ||||||||||||||||||||
Db       1108 GAAGCTGAACCACCCAATGTGCTGGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 1167

Qy        721 ATGATGATGCTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||  |||| ||||||||  |||||||
Db       1168 ATGATGATGTCCCCTCACTAAGCTGGCGGACAAGGAACTGGTGCACATGATTGGCTGGGCC 1227

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||| |||||||| |||||||| || |||
Db       1228 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGGCTCTTAGAAAGC 1287

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| ||||| |||| |||||||||||||| || ||||||||||||||
Db       1288 TGCTGGATGGAGGTGCTAATGGTGGGACTGATGTGGCGCTCCATCGACCACCCCGGCAAG 1347

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT 960
              ||||| || |||| || || |||||||||||||||||||||| |||||||| |||
Db       1348 CTCATTTTCGCTCCCGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTAGAAGGATT 1407

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||||| || || |||||||| || ||||||||||||||
Db       1408 CTGGAAATCTTTGACATGCTCCTGGCGACGACGTCAAGGTTCCGTGAGTTAAAACTCCAG 1467

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              ||||| || |||||||||||| |||||||||||||| ||||| |||||||||||||| ||
Db       1468 CACAAGGAGTATCTCTGTGTGAAGGCCATGATCCTCCTCAACTCCAGTATGTACCCTTG 1527

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |   | || | |||||| || || || ||||||||||||| | ||| | ||||||| |||
Db       1528 GCTTCTGCAAACCAGGAGGCAGAAAGTAGCCGGAAGCTGACACACCTACTGAACGCGGTG 1587

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||||| ||||| || ||||||||||||||||||| ||  |
Db       1588 ACAGATGCCCTGGTCTGGGTGATTGCGAAGAGTGGTATCTCCTCCCAGCAGCAGTCAGTC 1647

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCA 1241
              || ||||| |||||||||||||| || || ||||||||||
Db       1648 CGACTGGCCAACCTCCTGATGCTTCTTTCTCACGTCAGGCA 1688


RESULT 46
MMU81451
LOCUS       MMU81451               2152 bp    mRNA    linear   ROD 01-APR 1997
DEFINITION  Mus musculus estrogen receptor beta (Esrb) mRNA, complete cds.
ACCESSION   U81451
VERSION     U81451.1  GI:1912467
KEYWORDS    nuclear receptor; steroid receptor; transcription factor; mouse
            chromosome 12.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2152)
  AUTHORS   Tremblay,G.B., Tremblay,A., Copeland,N.G., Gilbert,D.J.,
            Jenkins,N.A., Labrie,F. and Giguere,V.
  TITLE     Cloninq, chromosomal localization, and tunctional analysis of the
            murine estroqen receptor beta
  JOURNAL   Mol. Endocrinol. 11 (4), 453-465 (1997)
  MEDLINE   97211383
   PUBMED   9058381
REFERENCE   2  (bases 1 to 2152)
  AUTHORS   Tremblay,G.B., Tremblay,A. and Giguere,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC 1996) Molecular Oncoloqy Group, Royal Victoria
            Hospital, 687 Pine Avenue West, Montreal H3A 1A1, Canada
```

```
FEATURES             Location/Qualifiers
     source          1. .2152
                     /organism="Mus musculus"
                     /strain="129/Sv"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /tissue_type="ovary"
     gene            1. .2152
                     /gene="Estrb"
     CDS             12   1469
                     /gene="Estrb"
                     /function="nuclear hormone receptor transcription factor;
                     steroid receptor"
                     /note="ERbeta"
                     /codon_start=1
                     /product="estrogen receptor beta"
                     /protein_id="AAB51132.1"
                     /db_xref="GI:1912468"
                     /translation="MTPYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPTSGHLSPL
                     ATHCQSSLLYAEPQKSPWCEARSLEHTLPVNPETLKRKLGGSGCASPVTSPSAKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACR
                     LRKCYEVGMVKGGSRRERCGYRIVRRQRSASFQVHCLNKAKRTSGHTPPVKELLINSI
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLLDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTARFRELKLQHKEYLCVKAMILLNSSMYPLATASQEAESSRKLTHLLNAVTDAL
                     VWVISKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHTLRGYKSSISGSECCSTEDSKSKEGSQNLQSQ"
BASE COUNT       549 a    526 c    570 g    507 t
ORIGIN

  Query Match             74.3%;  Score 929;  DB 10;  Length 2152;
  Best Local Similarity   84.3%;  Pred. No. 1.3e-199;
  Matches 1046;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||||||||    |  |||| ||||||||||||||| ||| |||||
Db         36 ATGAAGTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGCCAGACT 95

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCCC 120
              |||||||||||||| | ||||||||  |||| ||||| ||||||||||     |||  ||
Db         96 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 155

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| | ||| |||| |||||||||||||||| |||||||||||||||| |||
Db        156 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 215

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              |||||||||| |||||||||||||||| |||| |||||| || | |||| | ||||
Db        216 GAACACACCTTGCCTGTAAACAGAGAGACCCTGAAGAGGAAGCTTGGCGGGAGCGGTTGT 275

Qy        241 GCCAGCCCTGTTACTGCTCACAGTTCAAAGAGGGATGCTCACTTCTGGCGTGTCTGCAGC 300
              |||||||||||||| ||||| |  | ||||||||||||||||||||| ||||||||
Db        276 GCCAGCCCTGTTACTGCCCAAGCGCCAAGAGGGATGCTCACTTCTGGCGCCCTCTGCAGT 335

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| |||| || |||| || || |||||||| ||||||||||||||||||||||||||
Db        336 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 395

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCCACGAATCAGTGTACCATA 420
              |||||||||||||||||||||||||| ||||| |||||||| || |||||||||| ||
Db        396 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACCATA 455

Qy        421 GATAAAAACCCCGGCAAGAGGTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| || |||||||||||||||||||||| ||||||||||||| |||
Db        456 GACAAGAACCGGCGTAAAAGCTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 515

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| ||||||| || |||||||||||  | || || |||||||||
Db        516 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA 575

Qy        541 AGTGCCGAGGACCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGGGGGGCAGGGGCC 600
              ||||||  ||||||| |||| ||       ||| ||||||||||   | || ||| | |||
Db        576 AGTGCCAGCGAGCAGGTGCATTGCCTGAAGAAAGCCAAGAGAACCAGTGGGCACACACCC 635

Qy        601 CGAGTGGGGAGCTGCTGCTGGAGCGCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| |||||| || | ||||| |||||||||| |||||||||| |||
Db        636 CGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCCGAGCAGCTGGTGCTCACCCTGCTG 695

Qy        661 GAGGCTGAGCCCCCCCATGCTGATCAGCGGCCCCAGTGCCCCTTTCACCGAGGCCTCC 720
              || |||||||| ||| |||||||  | ||||| |||||   ||||||||||||||||||
Db        696 GAAGCTGAGCCACCCAATGTGCTAGTGAGCCGTCCCAGCATGCCCTTCACCGAGGCCTCC 755

Qy        721 ATGATGATGTCCCTCACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAACTGGGCC 780
              |||||||||||| || ||| |||| |||||||  |||| |||||||  |||||||
Db        756 ATGATGATGTCCCTCACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 815

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||||| |||||||| || |||||||| |||
Db        816 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC 875

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCCAATTGACCACCCCGGCAAG 900
              || |||||||||||| | ||| ||||||||||||||||||| || ||||||||||||||
Db        876 TGCTGGATGGAGGTGCTGATGATGGGGCTGATGTGGCGCTCCATCGACCACACCCGGCAAG 935

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              |||||||||||||||| || |||||||||||||||||||||||| ||||| ||||| |||
Db        936 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGATGAGGGCAAATGCGTGGAAGGGATT 995

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTGAGAGTTAAAACTGCAA 1020
              |||||||||||| |||||||||| || ||  || |||| || |||||||||| ||
Db        996 CTGGAAATCTTTGGCATGCTCCTGGCAACGACGGCAGGTTCCGTGAGTTAAAACTGCAG 1055

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||||||||| ||||| |||||||||| || ||||| |||||||||||||  ||
Db       1056 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTGCTCAACTCCAGTATGTACCCTTG 1115

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGCAAGCTGGCTCACTTGCTGAACGCGGTG 1140
              |  || || | |||||| || || ||||||||||| | ||| |  ||||||| |||
Db       1116 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1175

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||| | |||| || ||||| |||||||||| ||  |
Db       1176 ACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTGGAATCTCTTCCCAGCAGCAGTCAGTC 1235

Qy       1201 CGCCTGGCTAACCTCCTCATGCTCCTCTCCCACGTCAGGCA 1241
              || ||||| |||||||||||||| || || || ||||||||
Db       1236 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGGCA 1276


RESULT 37
A61588
LOCUS       A61588                1458 bp    DNA     linear   PAT 09-MAR-1998
DEFINITION  Sequence 6 from Patent WO9709348.
ACCESSION   A61588
VERSION     A61588.1  GI:3715918
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Kuiper,G G., Enmark,E. and Gustafsson,J.
  TITLE     ORPHAN RECEPTOR
  JOURNAL   Patent: WO 9709348-A 6 13-MAR-1997;
            KAROBIO AB (SE)
FEATURES             Location/Qualifiers
     source          1. .1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT       365 a    373 c    403 g    317 t
ORIGIN

  Query Match             73.7%;  Score 922.6;  DB 6;  Length 1458;
  Best Local Similarity   84.0%;  Pred. No. 3.6e-198;
  Matches 1042;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC  60
              ||||| |||||  |||||||||     |  |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGCCAGACT  84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC  120
               |||||||||||||| | ||||||||  |||| ||||| ||||||||||    ||| ||
Db         85 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC  144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA  180
              || | ||| | ||| ||||| ||||||||||||||||| ||||||||||||||||| |||
Db        145 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA  204

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC  240
              |||||||||| |||||||||||||||||| ||||| |||||| || | |||| | ||||
Db        205 GAACACACCTTGCCTGTAAACAGAGAGACCCTCAAGAGGAAGCTTGCCGGGAGCCGTTGT  264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC  300
              |||||||||||||||| |||| |  | ||||||||||||||||||||| || ||||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGATGCTCACTTCTGTGCCGTCTGCAGT  324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT  360
              | |||| |||| || |||| || || |||||||  ||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT  384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC  420
              |||||||||||||||||||||||||| |||| |||||||| || ||||||||||| ||
Db        385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA  444

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA  480
              || || |||||||| || | ||||||||||||||||||||| ||||||||||||||| |||
Db        445 GACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA  504

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA  540
              ||||| |||||||| ||| ||||||| || ||||||||||  | || || |||||||||
Db        505 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA  564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC  600
              ||||||  |||||||  |||| ||       ||| ||||||||||   | || ||| | |||
Db        565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCAGTGGGCACACACCC  624

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG  660
              || |||  |||||| ||||| || | |||| ||||||||||| ||||||||||| |||||
Db        625 CGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCCGAGCAGCTGGTGCTCACCCTGCTG  684

Qy        661 GAGGCTGAGCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC  720
              || |||||||| ||| |||||||  | || || ||||  ||||||||||||||||||||
Db        685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC  744

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC  780
              |||||||||||||| || ||| |||| |||||||  |||| ||||||||  ||||||||
Db        745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC  804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC  840
              ||||| || || |||||||||||||||||||| |||||||| || ||||||||||| |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC  864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTCGCGCTCAATTCACCACCCCGGCAAG  900
              || |||||||||||| | ||| ||||||||||||||||||| || ||||||||||||||
Db        865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTCGCGCTCCATCGACCACCCCGGCAAG  924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGGAAATGCGTAGAAGGAATT  960
              |||||||||||||||| || ||||||||||||||||||||||| |||| |||| |||
Db        925 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGATGAGGGGAAGTGCGTGGAAGGGATT  984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA  1020
              |||||||||||||||||||||||||| || ||  || |||| || |||||||||||||||
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTGCAG  1044

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG  1080
              |||||||||||||| |||| ||||||||||| || |||| ||||||||||||   ||
Db       1045 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG  1104

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTC  1140
              |  || || | |||||| || || || ||||||||||| | ||| |  |||||| |||
Db       1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG  1164

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG  1200
              || |||||  |||| |||||||| | ||||  | |||| ||||||||||| ||  |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGAATCTCTTCCCAGCAGCAGTCAGTC  1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCA  1241
              || |||||  |||||||||||||| || || || ||||||||
Db       1225 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGGCA  1265


RESULT 38
AR075922
LOCUS       AR075922                1458 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 6 from patent US 5958710.
ACCESSION   AR075922
VERSION     AR075922.1  GI:10002668
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Kuiper,G., Enmark,E.L.K. and Gustafsson,J.-A.
  TITLE     Orphan receptor
  JOURNAL   Patent: US 5958710-A 6 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="unknown"
BASE COUNT      365 a    373 c    403 g    317 t
ORIGIN

  Query Match           73.7%;  Score 922.6;  DB 6;  Length 1458;
  Best Local Similarity   84.0%;  Pred. No. 3.6e-198;
  Matches 1042;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC  60
              ||||| |||||  |||||||||     |  |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGCCAGACT  84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC  120
               |||||||||||||| | ||||||||  |||| ||||| ||||||||||    ||| ||
Db         85 GCAAGCCCAAATGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC  144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA  180
              || | ||| | ||| ||||| ||||||||||||||||| ||||||||||||||||| |||
Db        145 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA  204

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC  240
              |||||||||| |||||||||||||||||| ||||| |||||| || | |||| | ||||
Db        205 GAACACACCTTGCCTGTAAACAGAGAGACCCTCAAGAGGAAGCTTGCCGGGAGCCGTTGT  264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC  300
              |||||||||||||||| |||| |  | ||||||||||||||||||||| || ||||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGATGCTCACTTCTGTGCCGTCTGCAGT  324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT  360
              | |||| |||| || |||| || || |||||||  ||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT  384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC  420
              |||||||||||||||||||||||||| |||| |||||||| || ||||||||||| ||
Db        385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA  444

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA  480
              || || |||||||| || | ||||||||||||||||||||| ||||||||||||||| |||
Db        445 GACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA  504

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA  540
              ||||| |||||||| ||| ||||||| || ||||||||||  | || || |||||||||
Db        505 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA  564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC  600
```

```
Db        565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCAGTGGGCACACACCC 624

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| |||||| || | ||||| ||||||||||| ||||||||||| |||
Db        625 CGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCCAGCAGCTGGTGCTCACCCTGCTG 684

Qy        661 GAGGCTGAGCCCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || |||||||| ||| |||||||  | || || |||||   ||||||||||||||||||
Db        685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||| |||| ||||||||  ||||||||
Db        745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||||| |||||||| || |||||||| |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC 864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| | ||| |||||||||||||||||| || ||||||||||||||||
Db        865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              |||||||||||||||| || |||||||||||||||||||||| ||||| ||||||| |||
Db        925 CTCATCTTTGCTCCAGACCTTGTTCTGGACAGGGATGAGGGCAAGTGCGTGGAAGGGATT 984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||| || ||  || |||| || ||||||||||| ||
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTGCAG 1044

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCACTTG 1080
              |||||||||||||| |||||  |||||||||| || |||||  |||||||||||||   ||
Db       1045 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG 1104

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |  || || | |||||| || || || ||||||||||| | ||| |  |||||| |||
Db       1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1164

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCAGCAGCAATCCAGC 1200
              || |||||  |||| |||||||| | |||||  | |||| ||||||||||| ||  |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGAGTAGAATGTCTTCCCAGCAGCACTCAGCC 1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCAGGTCAGGCA 1241
              || ||||| |||||||||||||| || || || |||||||
Db       1225 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGCCA 1265


RESULT 39
AX022460
LOCUS       AX022460               1458 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 6 from Patent EP0935000
ACCESSION   AX022460
VERSION     AX022460.1  GI:10046870
KEYWORDS    .
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT      365 a    373 c    403 g    317 t
ORIGIN

  Query Match           73.7%;  Score 922.6;  DB 6;  Length 1458;
  Best Local Similarity   84.0%;  Pred. No. 3.6e-198;
  Matches 1042;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| |||||  |||||||||    | |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGGCAGACT 84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACCTTTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | ||||||||  |||| |||||  ||||||||||     |||  ||
Db         85 GCAAGCCCAAAFGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 144

Qy        121 CAGTTATCACATCTGTATGCGGAAGCTCAAAAGAGTCCCTCGTGTGAAGCAAGATCGCTA 180
              || | ||| | ||| ||||| ||||||||||||||||| ||||||||||||||||| |||
Db        145 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCCTTGC 240
              |||||||||||| ||||||||||||||||| ||||| ||||| || | |||| | ||||
Db        205 GAACACACCTTGCCTGTAAACAGAGAGACCCTGAAGAGGAAGCTTGCGCGGAGCGCTTGT 264

Qy        241 GCCAGCCCTGTTACTAGTCCAAGCACTAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              |||||||||||||||| |||| |  | ||||||||||||||||||| || ||||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAGCACTAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGT 324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| ||||| || |||| || || |||||||| ||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||| ||||| |||||||| || ||||||||||| ||
Db        385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA 444

Qy        421 GATAAAAACGGGGGCAACAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| || | ||||||||||||||||||||| ||||||||||||| |||
Db        445 GACAAGAACGGGGCTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 504

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| ||||||| ||| ||||||| || ||||||||||  | || || ||||||||
Db        505 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA 564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||  ||||||| ||| ||      ||| ||||||||||  | || ||| | |||
Db        565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCAGTGGGCACACACCC 624

Qy        601 CGAGTGCGGGAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| |||||| || | ||||| |||||||||| |||||||||| |||
Db        625 CGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCCAGCAGCTGGTGCTCACCCTGCTG 684

Qy        661 GAGGCTGAGCCCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || |||||||| ||| |||||||  | || || |||||   ||||||||||||||||||
Db        685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy        721 ATGATGATGTCCCTGACCAAGTTGGCCGACAAGGAGTTGGTACACATGATCAGCTGGGCC 780
              |||||||||||||| || ||| |||| |||||||| |||| ||||||||  ||||||||
Db        745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||||| |||||||| || |||||||| |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC 864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| | ||| |||||||||||||||||| || ||||||||||||||||
Db        865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCTCCATCGACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              |||||||||||||||| || |||||||||||||||||||||| ||||| ||||||| |||
Db        925 CTCATCTTTGCTCCAGACCTTGTTCTGGACAGGGATGAGGGCAAGTGCGTGGAAGGGATT 984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCGAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||| || ||  || |||| || ||||||||||| ||
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGACGGCACGGTTCCGTGAGTTAAAACTGCAG 1044

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCACTTG 1080
              |||||||||||||| |||||  |||||||||| || |||||  |||||||||||||   ||
Db       1045 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG 1104

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |  || || | |||||| || || || ||||||||||| | ||| |  |||||| |||
Db       1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1164
```

```
Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||| | |||||  | ||||| |||||||||||| || |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGACTAGAATCTCTTCCCAGCAGCAGTCAGTC 1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCA 1241
              || ||||| |||||||||||||| || || || ||||||||
Db       1225 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGGCA 1265


RESULT 40
MMAJ220
LOCUS       MMAJ220                 1458 bp    DNA     linear   ROD 30-NOV-1997
DEFINITION  Mus musculus mRNA for estrogen receptor subtype beta.
ACCESSION   AJ000220
VERSION     AJ000220.1  GI:2370296
KEYWORDS    estrogen receptor; subtype beta.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1458)
  AUTHORS   Pettersson,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1997) Pettersson K., Department of Medical
            Nutrition, Karolinska Institute, P 60 Novum, S-141 86 Huddinge,
            SWEDEN
REFERENCE   2  (bases 1 to 1458)
  AUTHORS   Pettersson K., Grandien,K., Kuiper,G.G. and Gustafsson,J.A.
  TITLE     The mouse estrogen receptor beta form estrogen response element
            binding heterodimers with the estrogen receptor alpha
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1..1458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             1..1458
                     /codon_start=1
                     /product="estrogen receptor, subtype beta"
                     /protein_id="CAA03949.1"
                     /db_xref="GI:2370297"
                     /db_xref="SPTREMBL:O35645"
                     /translation="MAFYSPAVMNYSVPSSTGNLEGGPVRQTASPNVLWPTSGHLSPL
                     ATHCQSSLLYAPPQKSPWCEARSLEHTLPVNPETLKPKTGGSGCASPVTSPSTKRDAH
                     FCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKNCQACR
                     LRKCYEVGMVKCGSHRERCGYRTVRRQRSASEQVHCLNKAKRTSGHTPRVKELLLNSL
                     SPEQLVLTLLEAEPPNVLVSRPSMPFTEASMMMSLTKLADKELVHMIGWAKKIPGFVE
                     LSLFDQVRLLESCWMEVLMVGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDM
                     LLATTARFRELKLQHKEYLCVKAMILLNSSMYPLVTATQSQEAESSEKLTHLLNAVTDAL
                     VWVIAKSGISSQQQSVRLANLLMLLSHVRHISNKGMEHLLSMKCKNVVPVYDLLLEML
                     NAHTLRGYKSSISGSQCSSTEDSKSKEGSQNLQSQ"
BASE COUNT       365 a    373 c    403 g    317 t
ORIGIN

  Query Match             73.7%;  Score 922.6;  DB 10;  Length 1458;
  Best Local Similarity   84.0%;  Pred. No. 3.6e-198;
  Matches 1042;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy          1 ATGAATTACAGCATTCCCAGCAATGTCACTAACTTGGAAGGTGGGCCTGGTCGGCAGACC 60
              ||||| ||||| |||||||||     | |||| ||||||||||||||| ||| |||||
Db         25 ATGAACTACAGTGTTCCCAGCAGCACCGGTAACCTGGAAGGTGGGCCTGTTCGCCAGACT 84

Qy         61 ACAAGCCCAAATGTGTTGTGGCCAACACCTGGGCACTTTCTCTCCTTTAGTGGTCCATCGC 120
              |||||||||||||| | |||||||    |||| |||| |||||||||(    ||| ||
Db         85 GCAAGCCCAAAIGTGCTATGGCCAACTTCTGGACACCTCTCTCCTTTAGCCACCCACTGC 144

Qy        121 CAGTTATCACATCTGTATGCGGAACCTCAAAAGAGTCCCTGGTGTGAAGCAAGATCGCTA 180
              || | ||| | ||| |||| ||||||||||||||||| ||||||||||||||||| |||
Db        145 CAATCATCGCTTCTCTATGCAGAACCTCAAAAGAGTCCTTGGTGTGAAGCAAGATCACTA 204

Qy        181 GAACACACCTTACCTGTAAACAGAGAGACACTGAAAAGGAAGGTTAGTGGGAACCGTTGC 240
              |||||||||| |||||||||||||||||| |||| |||||| || | |||| | ||||
Db        205 GAACACACCTTGCCTGTAAACAGAGAGACCCTGAAGAGGAAGCTTGGCGGGAGCGGTTGT 264

Qy        241 GCCAGCCCTGTTACTGGTCCAGGTTCAAAGAGGGATGCTCACTTCTGCGCTGTCTGCAGC 300
              ||||||||||||||| |||| |  | |||||||||||||||||||| || ||||||||
Db        265 GCCAGCCCTGTTACTAGTCCAAGCACCAAGAGGGATGCTCACTTCTGTGCCGTCTGCAGT 324

Qy        301 GATTACGCATCGGGATATCACTATGGAGTCTGGTCGTGTGAAGGATGTAAGGCCTTTTTT 360
              ||||| |||| || |||| || || |||||||| ||||||||||||||||||||||||
Db        325 GATTATGCATCTGGGTATCATTACGGTGTCTGGTCCTGTGAAGGATGTAAGGCCTTTTTT 384

Qy        361 AAAAGAAGCATTCAAGGACATAATGATTATATTTGTCCAGCTACAAATCAGTGTACAATC 420
              |||||||||||||||||||||||||||| |||| ||||||| || ||||||||||| ||
Db        385 AAAAGAAGCATTCAAGGACATAATGACTATATCTGTCCAGCCACGAATCAGTGTACGATA 444

Qy        421 GATAAAAACCGGCGCAAGAGCTGCCAGGCCTGCCGACTTCGGAAGTGTTACGAAGTGGGA 480
              || || |||||||| || | |||||||||||||||||||| |||||||||||| |||
Db        445 GACAAGAACCGGCGTAAAAACTGCCAGGCCTGCCGACTTCGCAAGTGTTACGAAGTAGGA 504

Qy        481 ATGGTGAAGTGTGGCTCCCGGAGAGAGAGATGTGGGTACCGCCTTGTGCGGAGACAGAGA 540
              ||||| |||||||| ||| ||||||| || |||||||||||  | || || |||||||||
Db        505 ATGGTCAAGTGTGGATCCAGGAGAGAAAGGTGTGGGTACCGAATAGTACGAAGACAGAGA 564

Qy        541 AGTGCCGACGAGCAGCTGCACTGTGCCGGCAAGGCCAAGAGAAGTGGCGGCCACGCGCCC 600
              ||||||  |||||||| |||| ||       ||| ||||||||||   | || ||| | |||
Db        565 AGTGCCAGCGAGCAGGTGCATTGCCTGAACAAAGCCAAGAGAACCAGTGGGCACACACCC 624

Qy        601 CGACTGCGGCAGCTGCTGCTGGACGCCCTGAGCCCCGAGCAGCTAGTGCTCACCCTCCTG 660
              || |||  |||||| |||||| || | |||| ||||||||||| ||||||||||| |||
Db        625 CGGGTGAAGGAGCTACTGCTGAACTCTCTGAGTCCCGAGCAGCTGGTGCTCACCCTGCTG 684

Qy        661 GAGGCTGAGCCCGCCCCATGTGCTGATCAGCCGCCCCAGTGCGCCCTTCACCGAGGCCTCC 720
              || |||||||| ||| |||||||  | || || ||||   ||||||||||||||||||||
Db        685 GAAGCTGAGCCACCCAATGTGCTAGTGAGTCGTCCCAGCATGCCCTTCACCGAGGCCTCC 744

Qy        721 ATCATCATCTCCCTCACCAAGTTGCCCGACAAGGAGTTGCTACACATCATCAGCTGCGCC 780
              |||||||||||||| || ||| |||| |||||||  |||| |||||||  |||||||||
Db        745 ATGATGATGTCCCTTACGAAGCTGGCTGACAAGGAACTGGTGCACATGATTGGCTGGGCC 804

Qy        781 AAGAAGATTCCCGGCTTTGTGGAGCTCAGCCTGTTCGACCAAGTGCGGCTCTTGGAGAGC 840
              ||||| || || |||||||||||||||||||||| |||||||| || |||||||| |||
Db        805 AAGAAAATCCCTGGCTTTGTGGAGCTCAGCCTGTTGGACCAAGTCCGCCTCTTGGAAAGC 864

Qy        841 TGTTGGATGGAGGTGTTAATGATGGGGCTGATGTGGCGCCTCAATTGACCACCCCGGCAAG 900
              || |||||||||||| | ||| ||||||||||||||||||| || ||||||||||||||
Db        865 TGCTGGATGGAGGTGCTGATGGTGGGGCTGATGTGGCGCCTCCATCCACCACCCCGGCAAG 924

Qy        901 CTCATCTTTGCTCCAGATCTTGTTCTGGACAGGGATGAGGGCAAATGCGTAGAAGGAATT 960
              ||||||||||||||| || |||||||||||||||||||||||| ||||| |||| |||
Db        925 CTCATCTTTGCTCCAGACCTCGTTCTGGACAGGGATGAGGGCAAGTGCGTGGAAGGGATT 984

Qy        961 CTGGAAATCTTTGACATGCTCCTGGCAACTACTTCAAGGTTTCCAGAGTTAAAACTCCAA 1020
              |||||||||||||||||||||||||||| || ||  || |||| || ||||||||||| ||
Db        985 CTGGAAATCTTTGACATGCTCCTGGCGACGAGGCACGGTTCCGTGAGTTAAAACTGCAG 1044

Qy       1021 CACAAAGAATATCTCTGTGTCAAGGCCATGATCCTGCTCAATTCCAGTATGTACCCTCTG 1080
              |||||||||||||| ||||| |||||||||| || |||| ||||||||||||    ||
Db       1045 CACAAAGAATATCTGTGTGTGAAGGCCATGATTCTCCTCAACTCCAGTATGTACCACTTG 1104

Qy       1081 GTCACAGCGACCCAGGATGCTGACAGCAGCCCGGAAGCTGGCTCACTTGCTGAACGCCGTG 1140
              |  || || | |||||  || || || || ||||||||||| | ||| |  |||||| |||
Db       1105 GCTACCGCAAGCCAGGAAGCAGAGAGTAGCCGGAAGCTGACACACCTATTGAACGCAGTG 1164

Qy       1141 ACCGATGCTTTGGTTTGGGTGATTGCCAAGAGCGGCATCTCCTCCCAGCAGCAATCCATG 1200
              || |||||  |||| |||||||||| | |||||  | ||||| |||||||||||| || |
Db       1165 ACAGATGCCCTGGTCTGGGTGATTTCGAAGACTAGAATCTCTTCCCAGCAGCAGTCAGTC 1224

Qy       1201 CGCCTGGCTAACCTCCTGATGCTCCTGTCCCACGTCAGGCA 1241
              || ||||| |||||||||||||| || || || ||||||||
Db       1225 CGTCTGGCCAACCTCCTGATGCTTCTTTCTCATGTCAGGCA 1265


Search completed: June 16, 2003, 07:38:52
Job time : 3319.22 secs
```